DOCUMENT RESUME

ED 039 890 LI 001 943

AUTHOR	Podell, Harold J.
TITLE	Foundations of Constructing a Marketing Data Base; Profitable Applications of the Computer to Marketing Management.
INSTITUTION	Leasco Systems and Research Corp., Bethesda, Md.
PUB DATE	30 Sep 68
NOTE	82p.; Paper given at American Management Association Briefing Session # 6310-04
EDRS PRICE	EDRS Price MF-$0.50 HC-$4.20
DESCRIPTORS	*Automation, *Computer Programs, *Data Bases, Information Systems, *Input Output Analysis, *Management Systems
IDENTIFIERS	*Management Information Systems, Marketing Data Base, MIS

ABSTRACT

An introduction into the foundations of constructing a marketing data base is presented for the systems and marketing executives who are familiar with basic computer technology methods. The techniques and concepts presented are now being implemented by major organizations in the development of Management Information Systems (MIS). A marketing data base in this presentation is considered to be a logical or conceptual data base which does not exist as a physical entity in one segment of the corporate MIS, but, rather, exists as a logical set of data that can be located by a software "definition table." Other data bases can utilize any appropriate elements from the corporate MIS. A definition table defines each data base within the MIS. The concept developed here relies upon the ability to utilize one common data base as several data bases through the process of software manipulation, this serves as the corporate MIS and is updated in a unified manner. The advantages are seen in terms of file maintenance, time and storage requirements. (NH)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

AMERICAN MANAGEMENT ASSOCIATION

Briefing Session # 6310-04

Profitable Applications of the Computer to Marketing Management

Concurrent Session: B

FOUNDATIONS OF CONSTRUCTING A MARKETING DATA BASE

By Harold J. Podell

September 30, 1968

LI 001943

Leasco Systems & Research Corporation
4833 Rugby Avenue
Bethesda, Maryland 20014

ACKNOWLEDGMENT

The production of this document was facilitated by Herbert Koller and James Kumnick, Leasco Systems & Research Corporation, Bethesda, Maryland. Their technical review and comments have been incorporated into the text.

Harold J. Podell
September 30, 1968

TABLE OF CONTENTS

		Page
I.	INTRODUCTION	1
II.	DATA BASE DESIGN	3
	A. DATA VERSUS INFORMATION	3
	B. CONCEPT OF DATA	6
	1. Definition	6
	2. Attributes	6
	3. Locations	6
	C. TYPICAL DATA SET DEVELOPMENT SEQUENCE	7
III.	DATA BASE CONCEPTS	15
	A. DATA BASE EXAMPLE	15
	B. DATA BASE RELATIONSHIPS	17
	C. INPUT/OUTPUT RELATIONSHIPS	20
	D. INFORMATION CENTERS	23
	1. Basic Elements	23
	2. Automatic Functions	26
	3. Time Sharing	27
IV.	DATA BASE MANAGEMENT	32
	A. DATA STANDARD DATA LANGUAGE	32
	1. Functions	32
	2. Vocabulary	36
	3. Maintenance	36
	B. MANAGEMENT INFORMATION SYSTEM (MIS) RELATIONSHIPS	36
	1. Decisions Versus Data Base	36
	2. On-line Versus Off-line	37
	3. Manual Versus Automated	37

		Page
V.	THE CORPORATE MIS	38
	A. MIS ELEMENTS AND FUNCTIONS	38
	B. TYPICAL DATA BASE FILE SET	41
	C. TYPICAL BUSINESS TRANSACTIONS	41
	D. FORCASTING EXAMPLE	45
VI.	MARKETING DATA BASE DIMENSIONS	47
	A. PREDICTION AND CORRECTION	47
	B. NEW PRODUCT DEVELOPMENT	47
	C. COMPETITIVE LEAD TIME	47
	D. DECISION MAKING	49
VII.	TYPICAL DATA BASE CONSTRUCTION AND SEARCH TECHNIQUES	50
	A. OVERVIEW	50
	B. LINKAGE	52
	C. DICTIONARY	54
	D. INVERTED INDEX FILE	57
	E. FILE MAINTENANCE	59
	F. FILE SEARCH	62
	1. Inverted File	62
	2. Direct Access	63
	3. Combined	65
	G. SUMMARY	73
VIII.	FIGURE REFERENCES	74
IX.	SOURCE LIST	77

I. INTRODUCTION

An introduction into the foundations of constructing a marketing data base is presented for the Systems and Marketing Executives who are familiar with basic computer technology methods. A marketing data base concept is developed in seven logical and interrelated steps. Illustrations are utilized as appropriate to develop the theme from data base design to the data base management functions, to specific marketing data base viewpoints and construction techniques.

This presentation is directed toward contemporary third-generation data base management systems design and is developed within a communications environment. Techniques and concepts are presented that are now being implemented by major organizations in the development of Management Information Systems (MIS).

A marketing data base is displayed as an integral part of the overall corporate decision-making oriented common data base for MIS. The interdependence with the corporate MIS is emphasized when discussing a marketing data base. A marketing data base in this presentation is considered to be a logical or conceptual data base which does not exist as a physical entity in one segment of the corporate MIS: Rather, the marketing data base exists as a logical set of data that can be located by a software "definition table."

Other data bases such as the financial data base and the inventory management data base can utilize any appropriate elements from the corporate MIS, including those in the marketing data base. A definition table defines each data base within the MIS data base. The concept developed in this presentation relies upon the ability to utilize one common data base as several data bases through the process of software manipulation; this central or common data base serves the corporate MIS and is updated in an unified manner.

The advantages are apparent in terms of file maintainence, time and storage requirements.

The ability to define data bases (such as the marketing, financial, and inventory management data bases) is available in software systems by utilizing methods such as definition and location tables. The additional utility of using software to define data bases is that their definitions can evolve without reprogramming; table parameters can be changed and any new data bases added without changing the data per se.

II. DATA BASE DESIGN

A. DATA VERSUS INFORMATION

One of the managerial control requirements of a well designed management information system MIS is provided the channeling of information to the appropriate MIS users in a timely and economic manner.

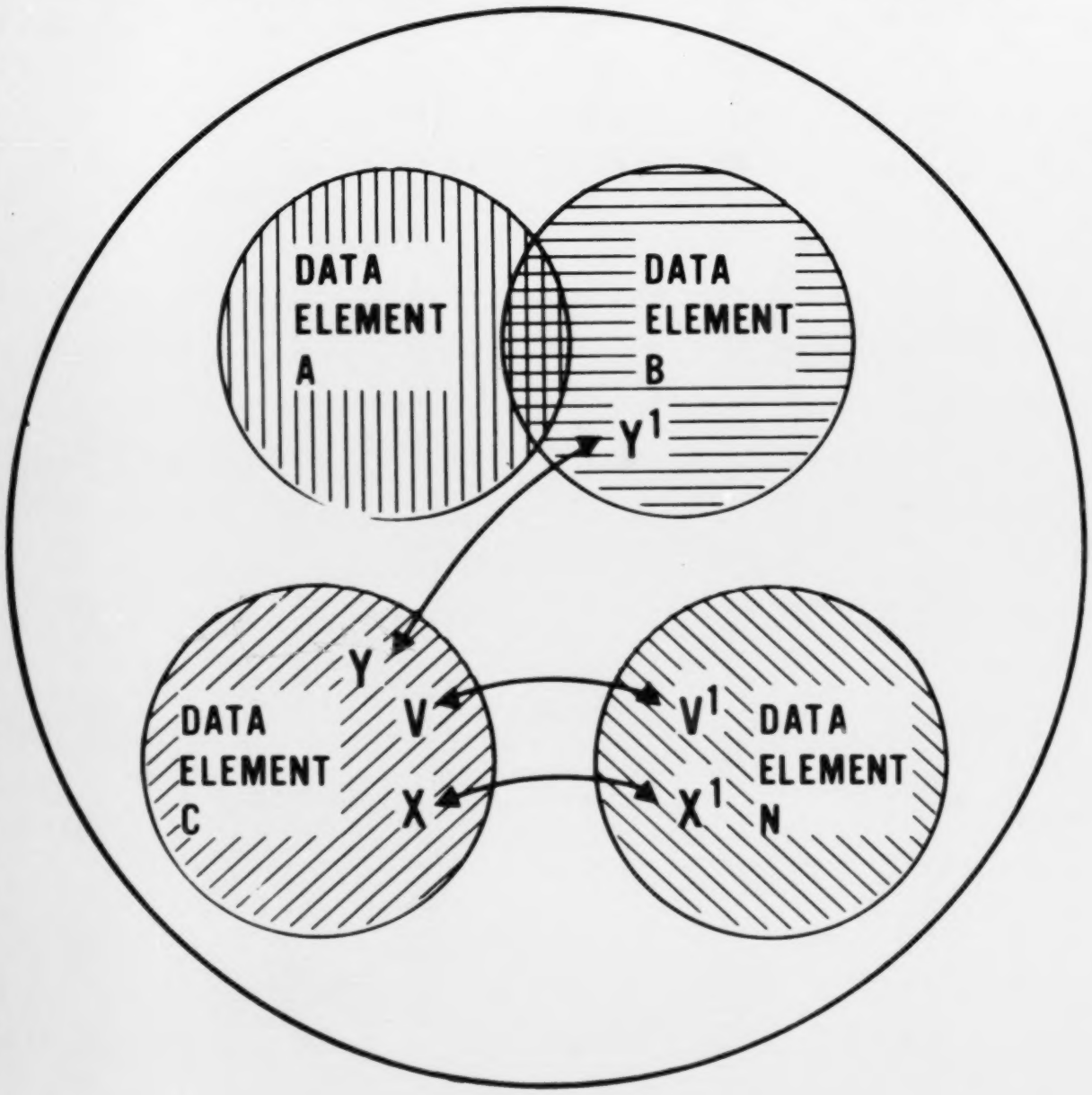
Data as defined in this presentation is a subelement of information and is contained in information. This relationship is outlined in the following figure. (Reference 1.)

Data elements as displayed to a manager in a management information system convey information that can be used in the information-decision process of an organization. It is the relationship of these data elements as shown in thus figure that is critical to effective design of a common data base system. In designing an information system, the data base configuration should provide for projected growth.

The marketing data base consists of those data in a common data base (as defined later in this presentation) that are directly concerned with the aspects of marketing in a given corporate MIS common data base. The specific definition of data elements varies from company to company. It is the objective of this presentation to illustrate the basic method for constructing a marketing data base within an overall common data base environment.

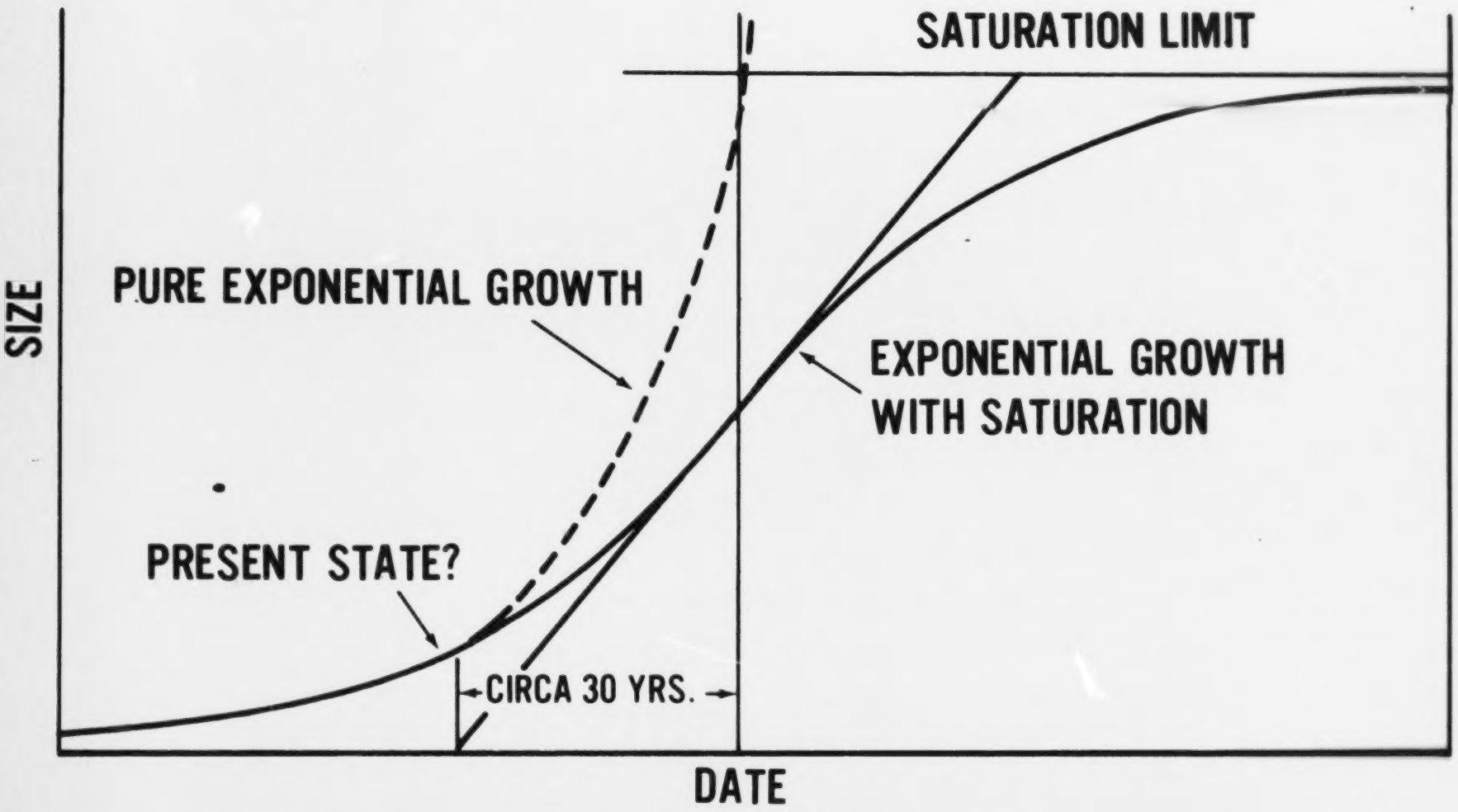
Data grows rapidly in the sense that a large volume of data such as marketing data is being generated both internally and externally within the corporate and economic environment. An example of the growth rate trend for data contained within information is illustrated in the following general logistics curve. This curve applies to the "S" shaped general form of the logistics growth curve and is equally applicable for data as well as information. (Reference 2.)

DATA IS CONTAINED IN INFORMATION



Reference 1.

GENERAL FORM OF THE LOGISTIC CURVE



B. CONCEPT OF DATA

1. Definition

There are many definitions for the word data. In the current context of constructing a marketing data base, data is considered to be those symbols which represent items. An item that is in the common data base will be defined as a marketing data base item if it meets the marketing data base criteria established by the marketing manager. Typical items that would meet these criteria are customer and product names.

2. Attributes

Symbols, which represent attributes of items, pertain to a description of defined data elements. For example, customer address, and telephone number; product description, price, cost and sales volume.

In a general sense, the definition facet of data can be considered the "noun" facet of data; and the attributes can be considered the "adjective" facet of data.

3. Locations

This last facet of the data description is the basic means of relating different data elements to each other by relative location in the common data base.

For example, the theme of this presentation is the construction of a marketing data base which, by its very nature, requires that each of the data elements in the common data base that are to be defined as data base elements must be located in a retrievable manner. The location approach usually consists of some type of linkage (as discussed later in this presentation).

In constructing a marketing data base, the method presented in this discussion utilizes a software table generation and maintenance system. This table system consists of a set of three types of tables such as:

- Definition tables
- Attribute tables
- Location tables

In more complex systems, these tables can be combined.

The important feature of software tables is that the elements that pertain to a marketing data base are (1) defined in an appropriate table, (2) their attributes are properly identified for each defined data element, and (3) that these marketing data base elements are related to each other by a type of software linkage.

An example of the linkage used is the development of a (3) location table is a table that lists each marketing data base element by a key number or mnemonic tag and cross references the tag with an appropriate address or pointer. Many techniques are utilized to define and construct a marketing data base within the table-oriented philosophy which has been outlined above. It is the purpose of this presentation to illustrate some of the more common concepts within this marketing data base construction philosophy and to relate the techniques to the MIS methodology.

C. TYPICAL DATA SET DEVELOPMENT SEQUENCE

In developing a data set such as a marketing data base, it is necessary to consider two basic dimensions of the data base elements:

1. The application of the data element within the MIS.
2. The computer software conventions and the design constraints.

In viewing applications of the data base element and data set element sequence, it is desirable to consider the total spectrum of applicable input/output data and information current and projected requirements for the MIS. In construction of the data base, there are a variety of procedures which can be followed. First, the basic mission of the particular MIS and the complete computer configuration including all software have to be defined. Given this complete set of information, a common data base and the constitutent data bases such as the marketing data base can be developed as follows (Ref: A typical data set procedure by Barbara Ambrose, Leasco Systems & Research Corporation.):

1. Define logical data set attributes
2. Define logical data elements attributes
3. Transcribe attributes to data definition form
4. Code table input
5. Generate system tables for data base
6. Code input data
7. Generate data set

The data base generation and maintenance procedures that are implied in this particular phase of the discussion are required for the use of generalized programs that are table-oriented and therefore flexible in handling of input data and in achieving file maintenance operations.

The first step of data set development is essentially defining the logical data set attributes, and these definitions are introduced via cards or equivalent. A sample data set format with appropriate attributes is presented in the next figure. (Reference 3.)

This sample data set is referenced by number and posted to the vendor number posting file.

Step 1. Define Logical Data Set Attributes

The first step in defining the data set attributes is shown in the above figure. The attributes in this case can be the number types of characters for fixed-length elements or for fields and so forth. A figure for an inverted file approach at the data-set level, which is also referred to as a record or set of fields.

In the inverted file, the data set is sequenced by vendor number in contrast to item number.

Step 2. Define Logical Data Elements Attributes

The second step consists of defining a logical data element classification of attributes and the data element format in this example is illustrated in the figure below. (Reference 5.)

SAMPLE SET FORMAT

THE FORMAT OF THE SAMPLE SET IS AS FOLLOWS:

ITEM NO.	VENDOR NO.	ITEM COST	NO. ITEMS	RECORDER CODE	ITEM DESCRIPTION	VENDOR NAME AND ADDRESS

ELEMENT NAME	ATTRIBUTES
ITEM NUMBER	6 NUMERIC CHARACTERS
VENDOR NUMBER	5 BYTE ALPHA-NUMERIC
ITEM COST	6 POSITIONS, 2 DECIMAL PLACES
NUMBER OF ITEMS	2 BYTE BINARY NUMBER
RECORDER CODE	1 BYTE NUMERIC CODE
ITEM DESCRIPTION	VARIABLE LENGTH, TEXTUAL
VENDOR NAME AND ADDRESS	VARIABLE LENGTH, TEXTUAL

Reference 3.

DATA DEFINITION TABLES

THE DATA SET TABLE FORMAT IS:

DATA SET ORGANIZATION CODES	DATA DEFINITION NAME	DATA SET NAME	DATA ELEMENT TABLE NAME

THE DATA ELEMENT TABLE FORMAT IS:

FIELD ID	DATA MODE	FIELD TYPE	FIELD POSITION	FIELD LENGTH	DECIMAL PLACES	POSTING FILE	POSTING FIELD	RECORD CRITERIA

Reference 5.

The data elements in this format could be considered the fields in a record. Each field has its own identifier as implied in the data element representation in the figure above.

The relationship of data definition tables at the data set or record level and the second level of detail that is the data element or data field level is shown in a typical illustration of the following figure. (Reference 6.)

A pointer titled DETSAMP at the data set level points to the data element level for the field DETSAMP which displays the applicable attribute definitions. This particular table follows a convention for the IBM 360/40 software constraints.

Step 3. Transcribe Attributes to the Data Set Definition Form

In establishing the attributes for use in a common data base, it is preferable to use the equivalent of parameter cards; this is part of the table-definition philosophy which is utilized in the entire file maintenance retrieval process for a large-scale MIS.

The advantages of table-oriented MIS systems are achieved in a rapid data base definition and flexible data base file maintenance and information retrieval process. The use of generalized programs which are capable of working with the various types of tables provides a minimum of reprogramming. A variety of attributes are identified for use in coding.

More advanced on-line systems can use terminals in lieu of these forms. The terminals would have keyboard and/or light pen control and computer feedback control to expedite data set definition as well as the entrie man/machine interactive process.

Step 4. Code Table Input

For the current example, step 4 is not illustrated. A sample coding for a data set and several supporting data elements, utilizing the format introduced above, could be introduced into the MIS using traditional keypunch or the newer off-line/or on-line keyboard systems. Data definition input-card equivalents are coded utilizing the data definition table formats which have been presented.

DATA DEFINITION TABLES

DATA SET TABLE

BQ	ORG	DD NAME	DS NAME	DET NAME
01	SQ	ANYDD	ANYDS	ANYDET
26	IS	SAMPLDS	SAMPLDS	DETSAMP
255	DI	ZZZZZDD	ZZZZZDS	ZZZZDET

DATA ELEMENT TABLE

DETSAMP

BQ	DATA MODE	FLD TYPE	POS	LGN	DEC PT	POST DS	POST DC	CRIT
03	1 ϕ	D ϕ	3	6	ϕ	ϕ	ϕ	E A
04	1 ϕ	7 ϕ	9	5	ϕ	112	101	E A
05	ϕ 2	7 ϕ	14	6	02	ϕ	ϕ	A 4
06	A ϕ	7 ϕ	20	2	ϕ	ϕ	ϕ	8 D
07	1 ϕ	7 ϕ	22	1	ϕ	ϕ	ϕ	ϕ 1
101	1 ϕ	1 ϕ	23	256	ϕ	ϕ	ϕ	ϕ ϕ
102	1 ϕ	1 ϕ	23	256	ϕ	ϕ	ϕ	ϕ 4
BIN	HEX	HEX	BIN	BIN	BIN	BIN	BIN	HEX

INTERNAL REPRESENTATION

Reference 6.

Step 5. Generate System Tables for Data Base

This program control phase can use a different type of table such as the location table which is also referred to as a system tag table. An example of a system tag table in the sample system is shown in the next figure. (Reference 8.)

Step 6. Code Input Data

Step 7. Generate Data Set

Steps 6 and 7 are not illustrated. These steps are used in creating a common data base as well as the constituent data bases, such as a marketing data base for the MIS. The coding of input data usually consists of introducing computer-coded data which is developed in a form that can be processed by the table-oriented system. This coding can be preceded by data conversion efforts if the MIS is a result of prior-integrated data base system. Generation of data sets is controlled by generalized programs that are table-oriented.

SYSTEM TAG TABLES

DATA SET TAG TABLE (MNEMONICS FOR EACH DATA SET IN SYSTEM)

DATA SET NAME	BQ	DETAG POINTER
ANYDATA	01	ANYSYST
SAMPLE1	26	DETAGPT
ZZZZZZZ	255	ZZZZZPT

DATA ELEMENT TAG TABLE FOR DATA SET

DATA ELEM, NAME	BQ
ITEM NO	03
VENDNOS	04
ITEMCST	05
NOITM	06
REOCODE	07
ITMDESC	101
NAMADR	102

III. DATA BASE CONCEPTS

A. DATA BASE EXAMPLE

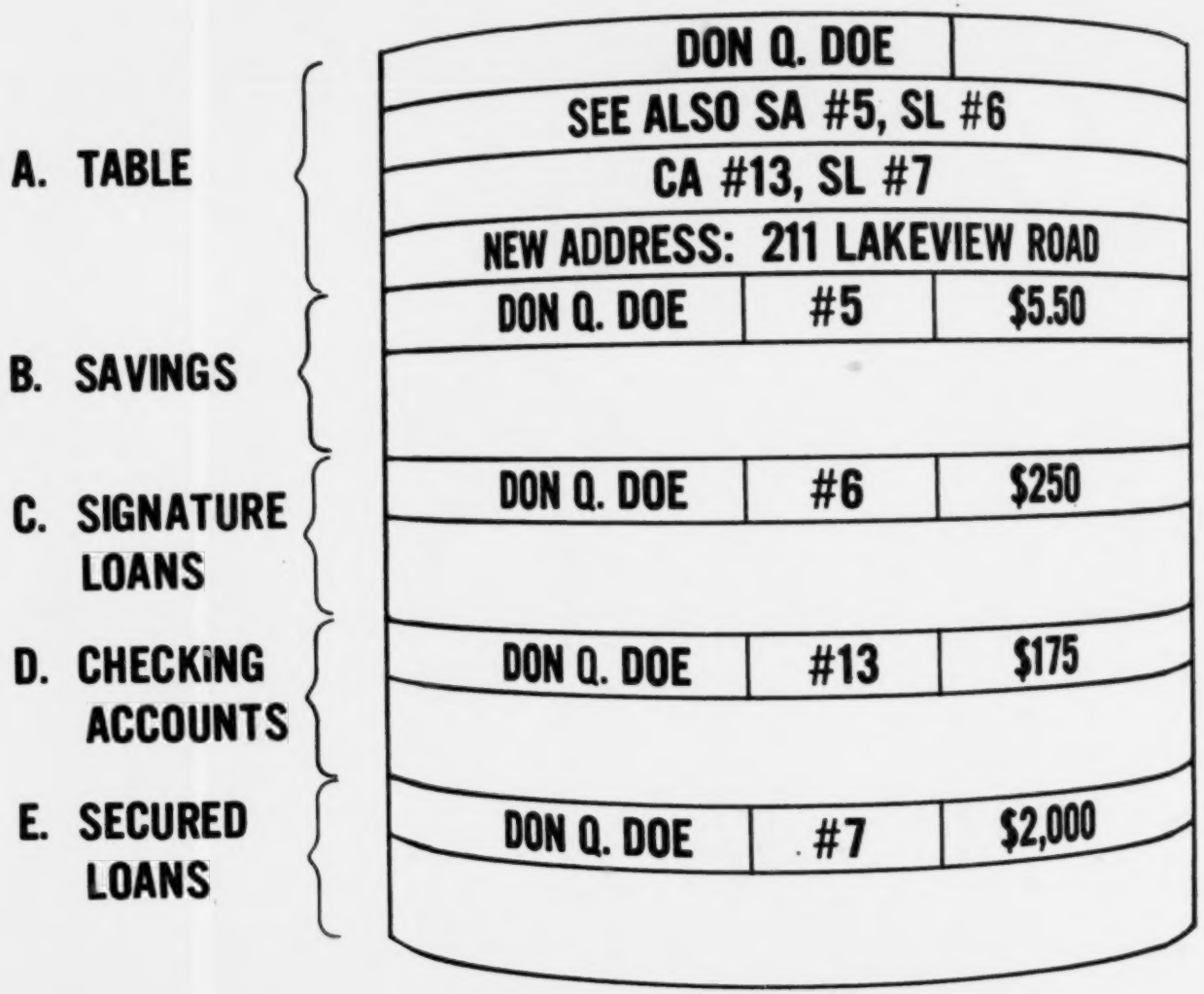
The data base concept illustrates the changing philosophies of data organization within a company. A simplified example of a segment of a financial data base within a common data base is presented in the figure below. (Reference 9.)

In a traditional bank, the accounts represented in this figure would each be separate entities recorded in the manual or manual/mechanized mode in an independent manner. The key here to historical data organization is the independent mode. If the set of data such as the savings information in a traditional information system is independent, considerable redundancy is implied in order to achieve independence.

This redundancy in the bank example could be illustrated by the location of the client address which is historically repeated for each of the four major accounts. That is, if an individual has a savings, signature loan, checking, and secured-loan account with one individual financial institution, then he could have four separate independent records under the historical data organization concept.

The figure shown above illustrates a more modern data base example where redundant information is stored in a definition and location table. When organizing a data base, file operation maintenance as well as the functional purpose of data should be investigated and evaluated. For example, in the preceding figure, an update to John Q. Doe's address for separate files (i.e. savings, signature, checking and secured loan) can be achieved with one change of a definiton table statement. This sample table also serves as a location table in that after the listing of John Q. Doe, references are made to locations that contain accounts that have records for John Q. Doe. This

A PARTIAL EXAMPLE OF DATA BASE CONSTRUCTION AND OPERATION



1. UPDATE TABLE STATEMENT

EXAMPLE:

CHANGE DON Q. DOE'S ADDRESS TO 211 LAKEVIEW ROAD

Reference 9.

same concept that was illustrated in the bank example applies to the development of a marketing data base.

B. DATA BASE RELATIONSHIPS

One of the most significant design relationship concepts that is associated with the data base is that of the conceptual constructural relationships as illustrated in the following figure. (Reference 10.)

As an illustration of a typical set of field characteristics, the following figure presents a managerial data base view of the field characteristics. (Reference 11.)

If this presentation is to have one single major point, this point is the recognition that a data base exists in two basic forms:

1. Conceptual
2. Structural

The conceptual data base is similar to the hardware configurator or a software data pool which defines in controlled vocabulary the source terms of all the elements within the common data base. This common data base defines the total MIS and selected elements within the conceptual data base are designated and table defined to constitute a marketing data base.

In the modern MIS philosophical approach, all of the decision-making functions within the organization are analyzed and the data and information required to perform these functions are identified in a data base configurator which is the conceptual MIS data base definition.

Within the conceptual data base definition for a given MIS, selected file sets can be identified for the entire MIS or for designated data bases such as a marketing data base. In general, the hierarchal relationship of the file sets is:

1. File set
2. File
3. Records
4. Field
5. Character or byte (not indicated in the above figures because it is usually not addressable.)

MANAGERIAL DATA BASE CONCEPTUAL/STRUCTURAL RELATIONSHIPS

- DECISION-MAKING FUNCTIONS (CONCEPTUAL)
  - DATA BASE CONFIGUATOR (CONCEPTUAL)
    - FILE SETS (STRUCTURAL)
      - FILES (STRUCTURAL)
        - RECORDS (STRUCTURAL)
          - FIELDS (STRUCTURAL)

Reference 10.

MANAGERIAL DATA BASE
TYPICAL FIELD CHARACTERISTICS

- ENGLISH TITLE AND MNEMONIC
- SECURITY CLASSIFICATION
  - INPUT/OUTPUT TAG
- FIELD SET IDENTIFIERS
  - VERTICAL (MASTER FIELD SET LINKS)
  - HORIZONTAL (PRIOR/NEXT DETAIL-FIELD LINKS)
- SORT KEYS

Reference 11.

This approach to the data base relationships is amenable to centralized control in a large MIS with local option at the structural level. The local information center selects from and participates in the development of the total managerial data base. Besides the many advantages mentioned for data base philosophies, this managerial data base concept utilizes controlled vocabulary and standard definition throughout large-scale MIS systems.

The elements of a managerial data base as presented can be generally related to the previous discussion concerning the concept of data:

1. Definition
2. Attributes
3. Locations

Addtional information which is shown in the figure above is related to security and sort keys as representative of special features required in various data management software systems (as discussed later in this presentation).

C. INPUT/OUTPUT RELATIONSHIPS

There are several ways to view the input/output relationships of an MIS. Since the marketing data base is table-defined within a MIS, the figures in this section are at the MIS detailed level with marketing data base subdetail implied.

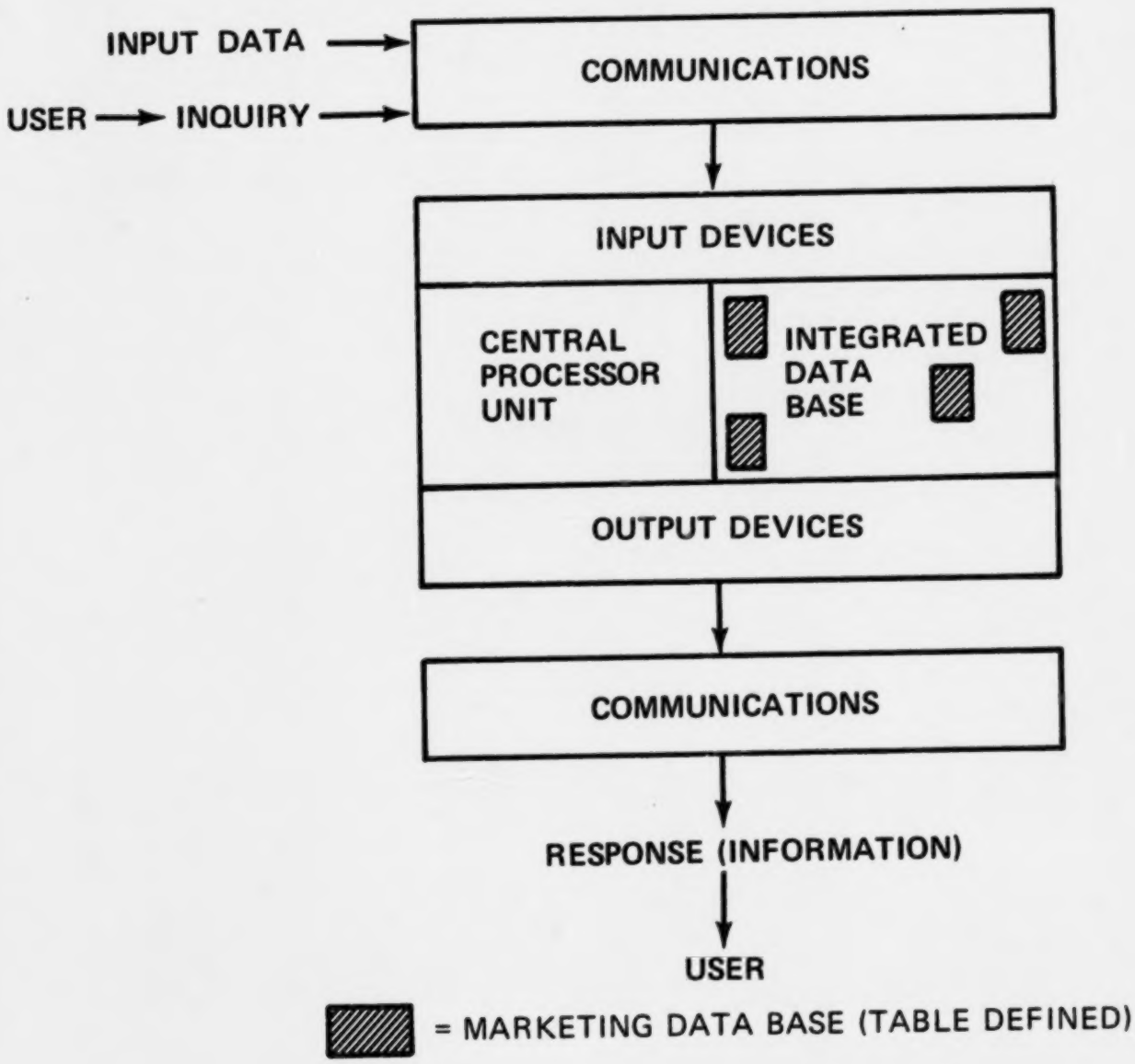
The first figure in the series indicates the relative location of the marketing data base elements within the MIS common data base to the integrated data base.

This relationship is representative of the basic concept of a data base and illustrates the conceptual/structural physical view inside an ADP storage device. (Reference 12.)

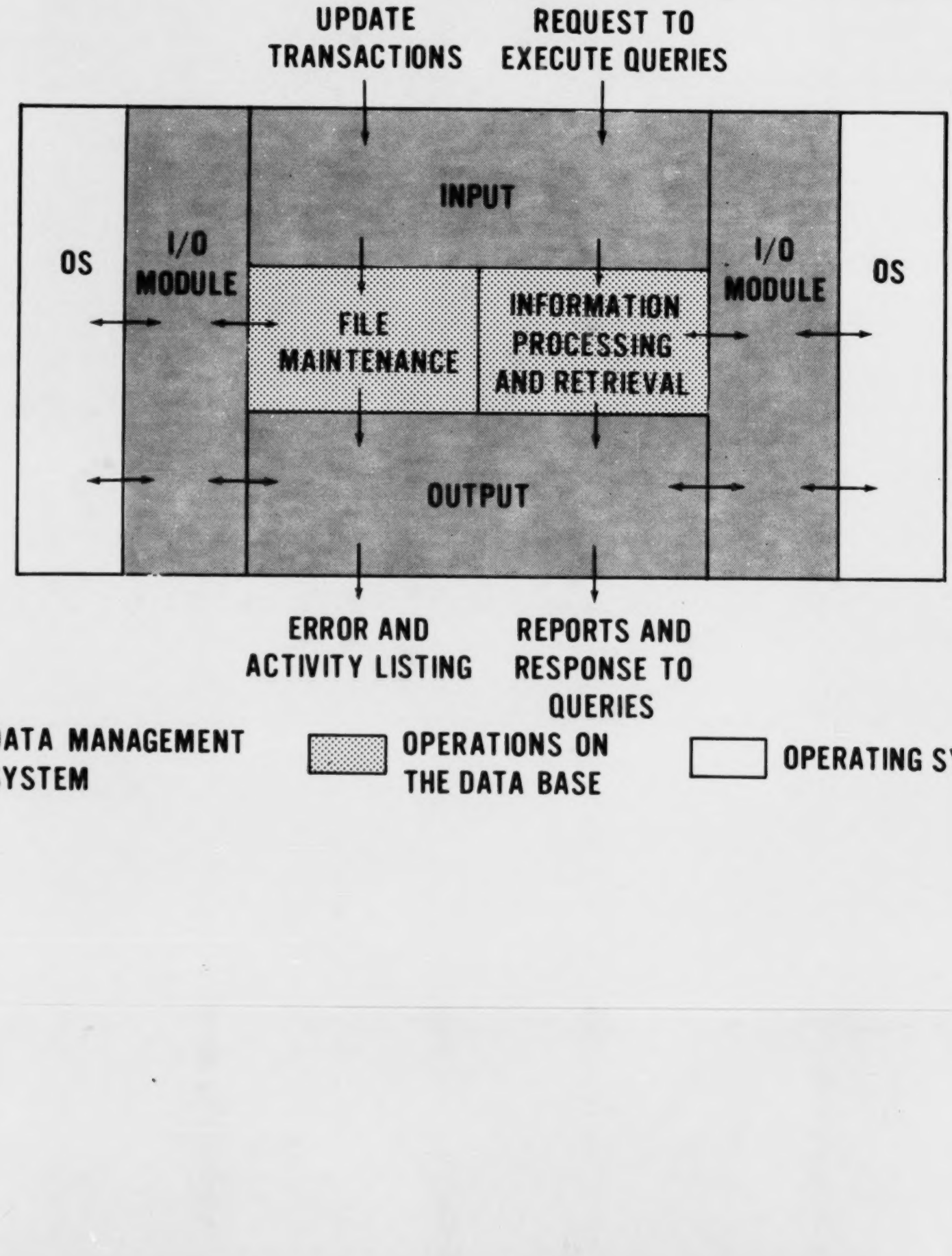
The figure below illustrates an overview of the data flow and the inquiry flow into an MIS system with appropriate response generated for the user. (Reference 13.)

The important feature of this modern approach to data base functioning is that the MIS common data base is totally contained within the hardware

THE DATA BASE CONCEPT



OVERALL VIEW OF A POSSIBLE DATA MANAGEMENT SYSTEM



configuration, and there is no direct communication with the elements of the data base: Rather, the user interfaces with a communication capability.

The communications capability presented in this discussion is that of a data base management system (as defined later). A basic overview of the logical control of the data management system over the file information retrieval operation on the common data base is illustrated in the above figure. Basic functions of a data base management software system as shown operating under an operating system are:

1. Update transaction
2. Request to execute queries
3. Error and activity listings
4. Reports and response to queries.

The input/output relationships that have been presented are concluded in the figure below concerning the data base concept. This figure illustrates the major software elements of a hardware/software information center operation for MIS. (Reference 14.)

The total information center is controlled by an operation system which integrates the hardware and software function. In the operating system, a data management language operates and utilizes various software capabilities, such as:

1. Tables (directory)
2. Application programs.

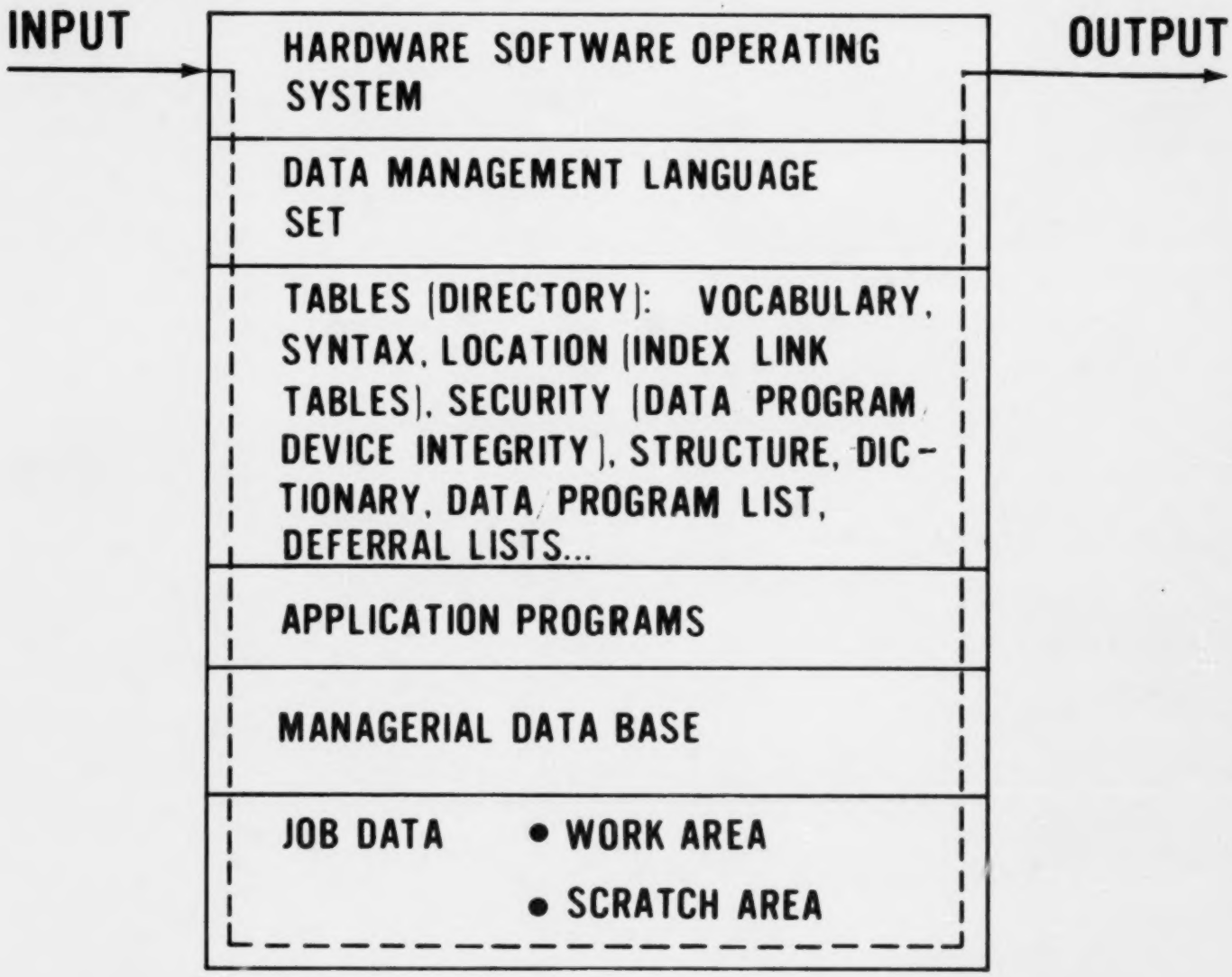
The application programs operate under the data management language functions, utilizing tables and the common data base. The job data are stored in process areas.

D. INFORMATION CENTERS

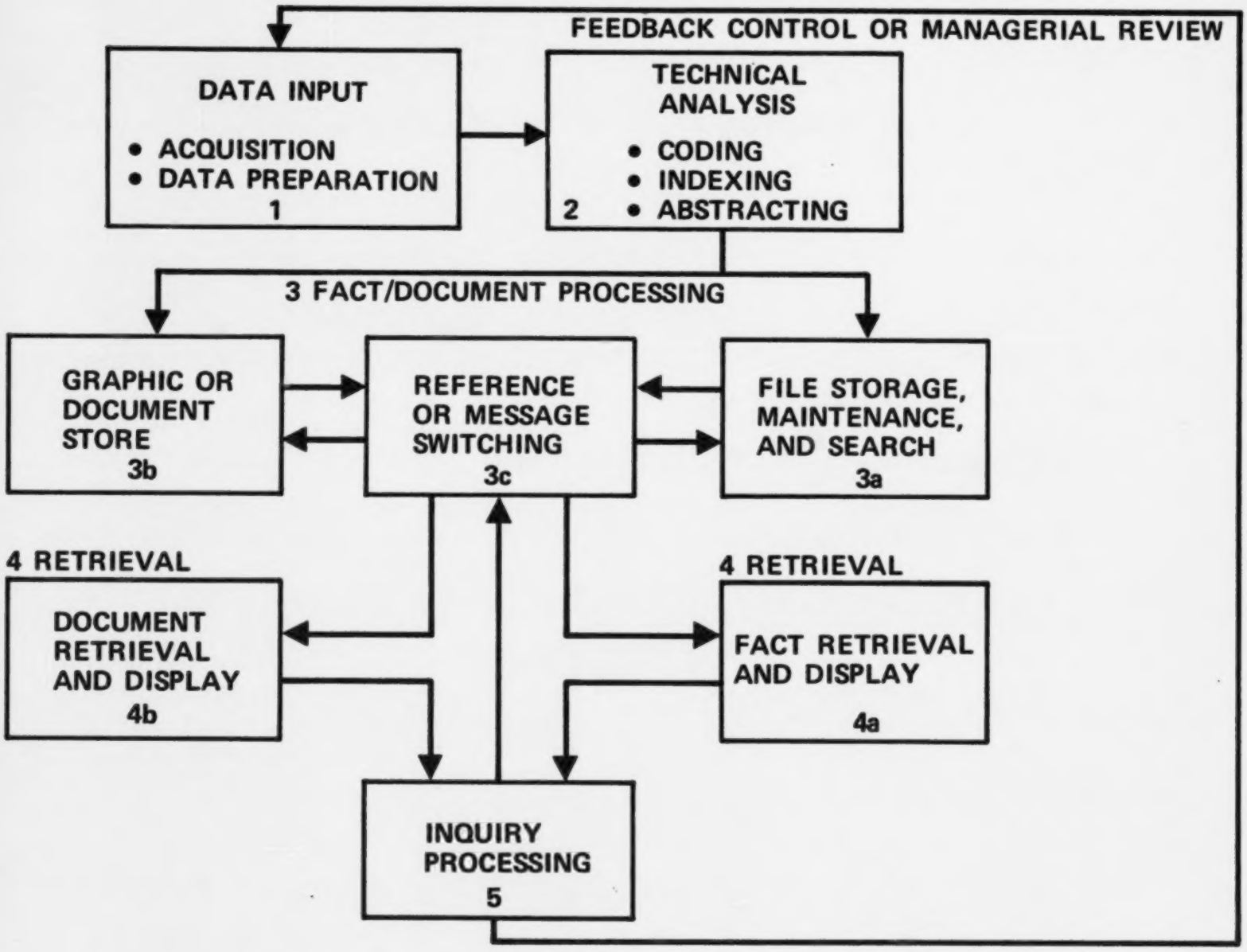
1. Basic Elements

A total information center hardware/software overview is presented below from a basic element or function viewpoint. This presentation is related to the preceding input/output relationships; Hardware/software interrelationships are not identified. (Reference 15.)

THE DATA BASE CONCEPT


INPUT
HARDWARE SOFTWARE OPERATING SYSTEM
OUTPUT
DATA MANAGEMENT LANGUAGE SET
TABLES (DIRECTORY): VOCABULARY, SYNTAX, LOCATION (INDEX LINK TABLES), SECURITY (DATA PROGRAM DEVICE INTEGRITY), STRUCTURE, DICTIONARY, DATA/PROGRAM LIST, DEFERRAL LISTS...
APPLICATION PROGRAMS
MANAGERIAL DATA BASE
JOB DATA
• WORK AREA
• SCRATCH AREA

ELEMENTS OF AN INFORMATION CENTER



An MIS provides a means for the manager to achieve exception reports and summary information on a top-management level and for middle or line management to perform basic operations that require processing information. The marketing data base serves staff management, as well as top management, in cutting across traditional line or functional organization within the organization and MIS data base, in order to produce meaningful and timely marketing reports and analyses.

In order to achieve this middle/staff/top management role, an information center can be designed to have computer-based MIS, as illustrated in the figure above. The man/machine interface is achieved in Block No. 5, Inquiry Processing. Both demand and periodic reports are developed and communicated to respective users through a series of events starting and ending in the inquiry processing block (5).

A comprehensive management information center mechanism is shown in the figure above: The right-hand side pertains to the storage, maintenance and search of facts or data, which is the most common element in current MIS systems. The left-hand side of the figure shows the interrelationships (usually through accession numbers) of the data in the computer system with appropriate graphic or source-document images. In the future, MIS will tend toward an interfacing of both the left-hand side of the figure (graphic or document information) with the right-hand side (fact or data processing functions).

The emphasis of this presentation is upon data processing and data base design for the marketing data base within the MIS, and is concerned, primarily, with the internal systems operation. The internal systems operation pertains to a data base of facts or data that are required to support the MIS and related data bases, e.g., marketing data base.

2. Automatic Functions

The trend today is toward automated information centers that provide MIS functions, as well as related information processing services, in a time-shared and multiprogramming mode. Communication media that assist the man/machine interface in an on-line manner are most common today in

terms of plans and of development. An abstract view of an automated information center is presented in the following figure. (Reference 16.) The key blocks of the automated information center that are related to the central theme of designing and constructing a marketing data base for MIS operation are 3-a, 3-c, and 4-a.

These blocks are connected with the file storage and maintenance operation in both main and auxiliary memory (3-a) and the complex message switching system which is under a joint operating/hardware control mechanism (3-c): These two internal operations, when coordinated through a fact retrieval and display function (4-a), support the necessary input/output processing mode in an MIS.

The marketing information storage and retrieval operations perform as a subset of the MIS inquiry processing mode.

3. Time Sharing

A data base, such as a marketing base, operates within an MIS system environment, which can include time-sharing functions. For reader convenience, several definitions are presented below:

(1) On-line systems are those systems in which the input/output terminals are under computer control.

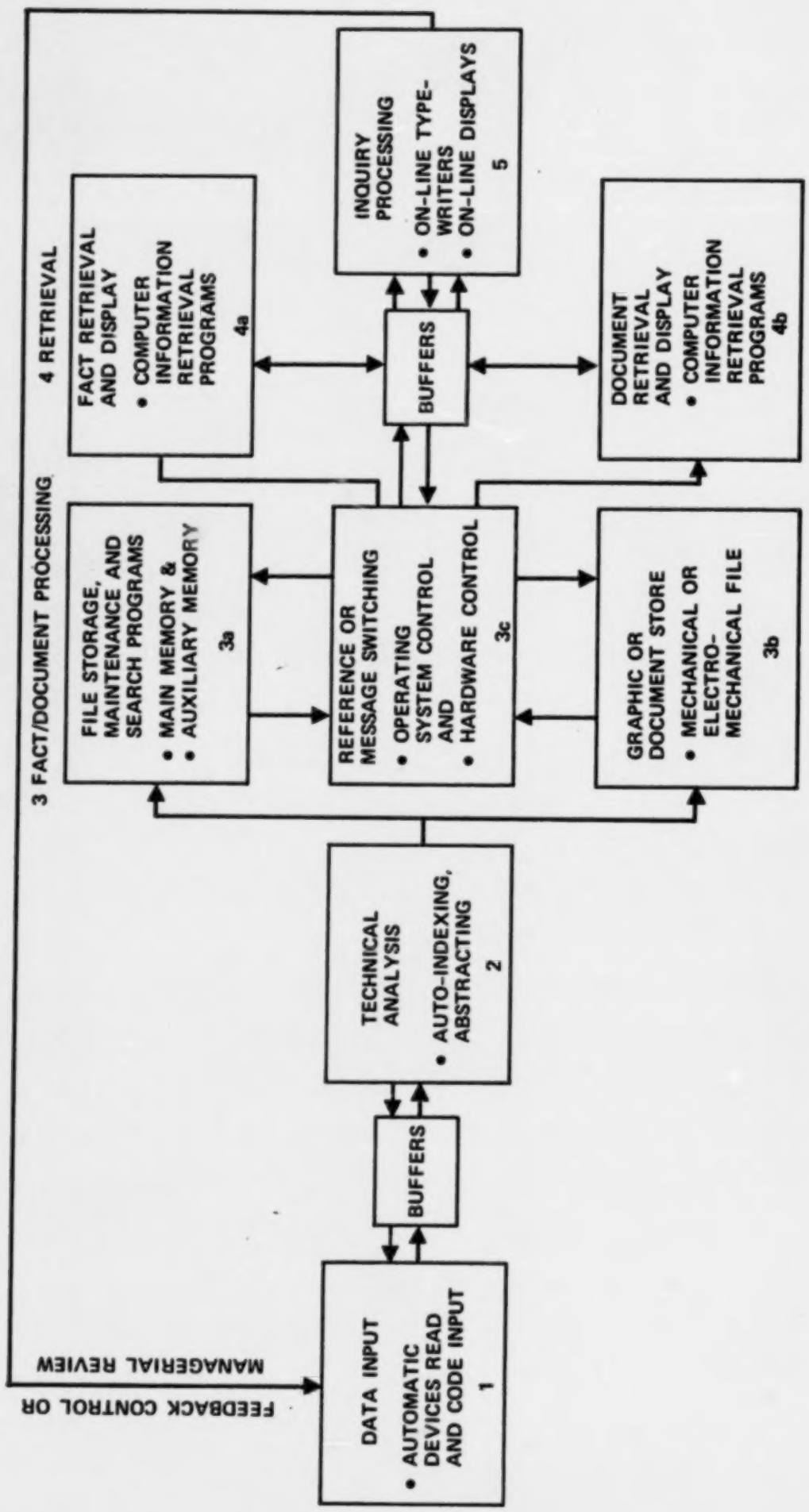
(2) Multiprogramming systems refer to a computer where two or more tasks can be performed in one system during a given time interval.

(3) Multiprocessing systems are those computer complexes that have more than one central processor under operating system control.

(4) Time-sharing systems are on-line computer networks which have multiprogramming and/or multiprocessing capabilities. In general, a terminal user in the network experiences little or no delay because of the ready availability of the base and the programs.

Reference 16.

AN AUTOMATED INFORMATION CENTER



(5) Real-time systems are closed-loop computer networks where data are introduced in a continuous or on-line mode and the decisions must be made in a relatively short time by man and/or machine. The significant characteristic of a real-time system is that there is little control of input/output rates so that the entire system design is to ensure no degradation. (Source: Jim Kumnick, Leasco, Bethesda, Maryland.)

A common data base system that is stored and operating in a digital computer consists of a set of alphanumeric data that represents the information process for the overall MIS and constituent table defined data bases, e.g. marketing data base. These data are stored in a computer configuration which is buffered with input/output devices and which communicates with a user in inquiry/response capability. The communications input/output relationships that have been introduced for an MIS imply a considerable degree of multi-programming or time sharing as illustrated in the following figure. (Reference 17.)

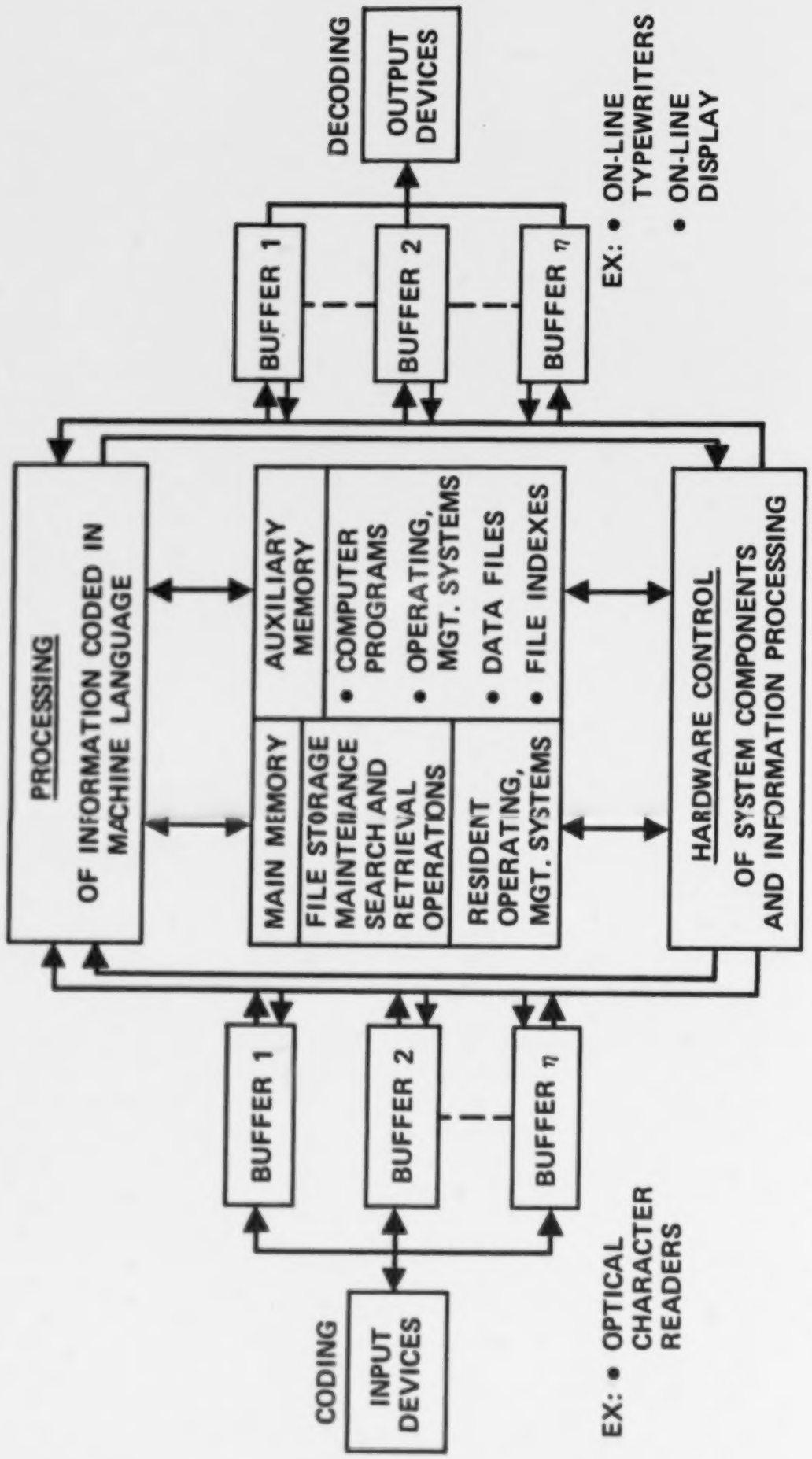
An advanced digital computer time-sharing and multiprogramming system, as envisioned today with consideration for future on-line trends is illustrated in the figure above. The common data base is indicated as a set of data files in auxiliary memory. The file indexes apply to the data file or data base system and will be resident with the data files in auxiliary memory. This is one advantage of the common data base, i.e., it is on-line to an MIS advanced system.

For purposes of discussion, the common data base which includes the table defined marketing data base, is considered resident in auxiliary memory. The programs that operate on the data base are stored in auxiliary memory (as shown in the figure above). The implication is that a marketing manager who has access to an input/output device can receive timely marketing information and analyses at his console or terminal.

The multiprocessing capability is relatively new in time-sharing systems but is projected to become more common in the future. Development

Reference 17.

AN ADVANCED DIGITAL COMPUTER TIME-SHARING SYSTEM



of time-sharing systems in an MIS environment will require consideration of data base design methods as presented in this discussion. Indexing, file maintenance, and inquiry processing methods that are compatible with the advanced hardware/software system capabilities should also be evaluated. Time-sharing systems development for MIS should also consider the projected growth and direction of corporate data trend. This applies to the total corporate data base as well as constituent data bases, e.g., marketing data base.

IV. DATA BASE MANAGEMENT

A. STANDARD DATA LANGUAGE

1. Functions

There are many data base management languages in various stages of development and operation. The total list of data base management languages or more commonly data management systems would exceed 25 basic offerings. These languages have varying degrees of operating system functions and general programming language interfaces. In general, the more common data management languages are being developed for IBM 360 systems, e.g. TDMS: time-shared data management system (SDC),GIS: Generalized Information System (IBM), and DM-1: Data management-1 (Auerbach).

Other systems are designed for non IBM computers such as IMRADS: Information Management, Retrieval, and Dissemination System (UNIVAC) and INFOL: Information Oriented Language (CDC).

There are many data base management system functions. Some of these functions are illustrated in the following figures. The functions listed in the figure are compared below with standard functions (Ref. Jack Minker and Jerome Sable, File Organization and Data Management: Chapter 5, ANNUAL REVIEW OF INFORMATION AND TECHNOLOGY,1967):

Figure	Minker and Sable
1. File Transactions - Historical	(1) Data description language and (2), (2a), (3).
2. File Processing - Business/ Scientific	(2) Job specification language and (2a) File maintenance language
3. File Input/Output Dialog - Inquiries	(3) Query language
4. Control Mode	(4) Operating System Functions

IMPLIED FUNCTIONS

(4a) System processing capability
(4b) Programmer interface

Although data management systems can have operating system functions, it is desirable to distinguish in general between the two types of software systems (ibid. Minker and Sable): (Reference 18.)

(1) Data management system functions

Operating under the operating systems "manages the access, manipulation, and protection of the logical items within the physical cells or blocks."

(2) Operating system functions

"Manages the allocation, movement, and protection of physical cells (blocks) among various levels of storage."

In general, operating systems are concerned with hardware/software job production functions and not with the contents of the physical cells or blocks that are being processed.

A common data base management file structure is illustrated by a purchasing data example in the following figure. (Reference 18a)

NOTE: The hierarchical relationship that is illustrated in DM-1 above differs from the general terminology introduced earlier in this discussion.

NOTE: In the DM-1 format, the letter "R" indicates that the item is a record which is repeated in the data base as many times as the item is recorded.

In general, a given data management system does not necessarily have all of the data base management functions that have been presented above.

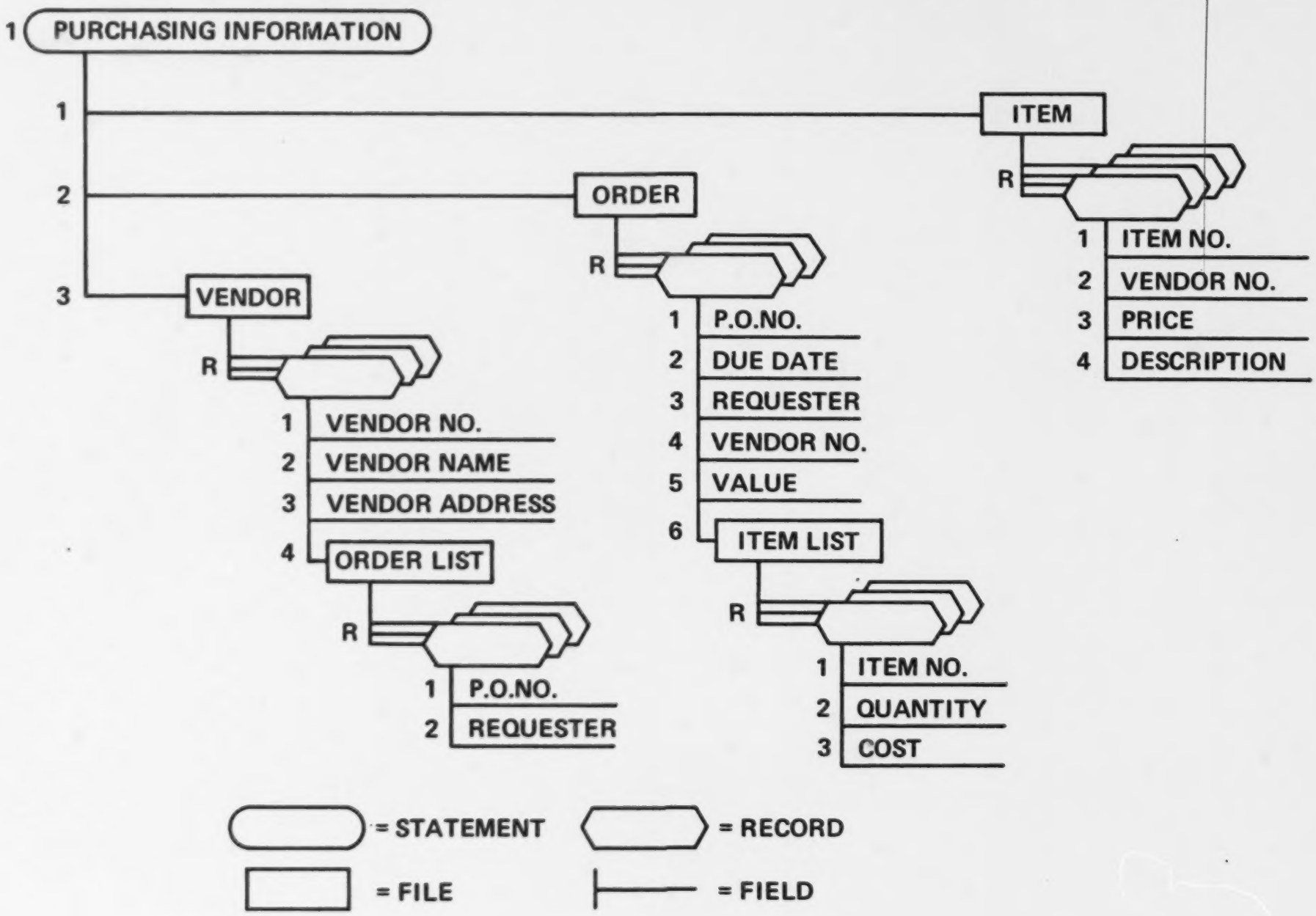
The categories given such as (1) file transactions-historical, (2) file processing-business/scientific, (3) file input/output dialog-inquiries and (4) control mode are interrelated. Depending upon the objectives and hardware/software configuration utilization, each corporate MIS and

DATA BASE MANAGEMENT TYPICAL LANGUAGE FUNCTIONS

- FILE TRANSACTIONS - HISTORICAL
  - INPUT
    - COMMUNICATIONS
    - QUEUEING
  - EDITING
  - MAINTENANCE (UPDATE, CONDITIONAL CHANGE, REORGANIZE, RESTRUCTURE, SAFEGUARD)
  - OUTPUT
- FILE PROCESSING - BUSINESS/SCIENTIFIC
  - MAINTENANCE
  - CALCULATION
  - SIMULATION
- FILE INPUT/OUTPUT DIALOG-INQUIRIES
  - COMMUNICATIONS
  - FACT RETRIEVAL
  - DOCUMENT RETRIEVAL
  - REPORT GENERATION
  - CONVERSATIONAL MODE
  - GRAPHIC I/O INTERACTION
- CONTROL MODE :
  - HARDWARE/SOFTWARE
  - OPERATING SYSTEM

Reference 18.

STRUCTURE OF PURCHASING DATA IN DM–1 : EXAMPLE



Reference 18a.

constituent data base system requires a different type of data base management language system. Commercial or specialized data management systems can be used. The selection is a function of system specifications, hardware configuration, and managerial objectives.

The important dimension of this development is that when constructing a marketing data base, it is necessary to construct a data base within the overall MIS and data management system constraints that apply to a particular organization. Marketing data bases in today's business environment are constructed in conjunction with the development of corporate information processing systems and relate to the marketing decision making process.

2. Vocabulary

The vocabulary in a data base management system is controlled as has been indicated above. The implication for constructing a marketing data base is that the marketing management should actively participate in the vocabulary development and standardization process for the corporate MIS. Many corporate functions involve marketing and the vocabulary concerned with typical corporate functions should be reviewed by marketing management to ensure the development of a meaningful reliable common and constituent data base system.

3. Maintenance

No data base system is static,and, as has been mentioned previously, an advantage of data base management systems is in the ability to maintain the data base in a uniform, rapid, effective cross reference manner.

B. MANAGEMENT INFORMATION SYSTEM (MIS) RELATIONSHIPS

1. Decision Versus Data Base

Since the data base designed and construction reflects the needs of the decision making process, it is vital that major marketing functions be adequately represented in the data base system, e.g. sales forecasting and projection, marketing research, advertising, strategic marketing, planning, and long-range planning functions.

2. On-Line Versus Off-Line

Within the decision making approach presented in this discussion, all corporate decision making operations should be represented in the conceptual data base design. This statement implies that regardless of whether processing is off-line or on-line, the vocabulary and related file maintenance methods should be developed for the corporate MIS and constituents such as a marketing data base.

3. Manual Versus Automated

The same type of reasoning applies to the method of operation of the MIS and the constituent data bases. Whether or not a data base is automated or manual in today's operation, its definition in relation to the decision making processes of the corporation should be included in the overall conceptual design of of the MIS data base.

V. THE CORPORATE MIS

A. MIS ELEMENTS AND FUNCTIONS

In a broad sense, an MIS consists of the following elements:

1. Business Operating System
   - (1) Data origination
   - (2) Data collection, preparation and input processing
   - (3) Data base maintenance and processing

2. Control System, Data Collection, Analysis
   - (1) Data collection, analysis
   - (2) Routine decisions
   - (3) Reports to management (a) periodic, (b) demand
   - (4) Managerial feedback control functions

The MIS functions parallel the information-decision process for a typical corporation. The information flow involves broadly the interaction of management plans and objectives of a controlled operating set of systems (business rather than a computer operating system.)

The control system functions requires a feedback of summary and exception reports to management as a part of the dynamic process of managing and planning for a successful business operation. In addition, data collection, analysis and routine decisions are generated within the control system.

It is implicit here that each system such as the control and operation systems are man-machine information-decision systems.

An MIS consists of a set of procedures that operates within a computer based configuration to assist in the flow of information, instructions, orders, and specifications within and between the control and operating systems for a business organization. The control system requires data collection analysis, storage procedures, file systems, and supporting functions. In a complementary manner, the business operating system also requires methods for storing and processing the information to interact with the control system.

The need for an MIS is at least a three level need. Level one consists of the top management level for summary and exception reports. The control system on level two serves as a buffer for top management when interfacing with the operating elements.

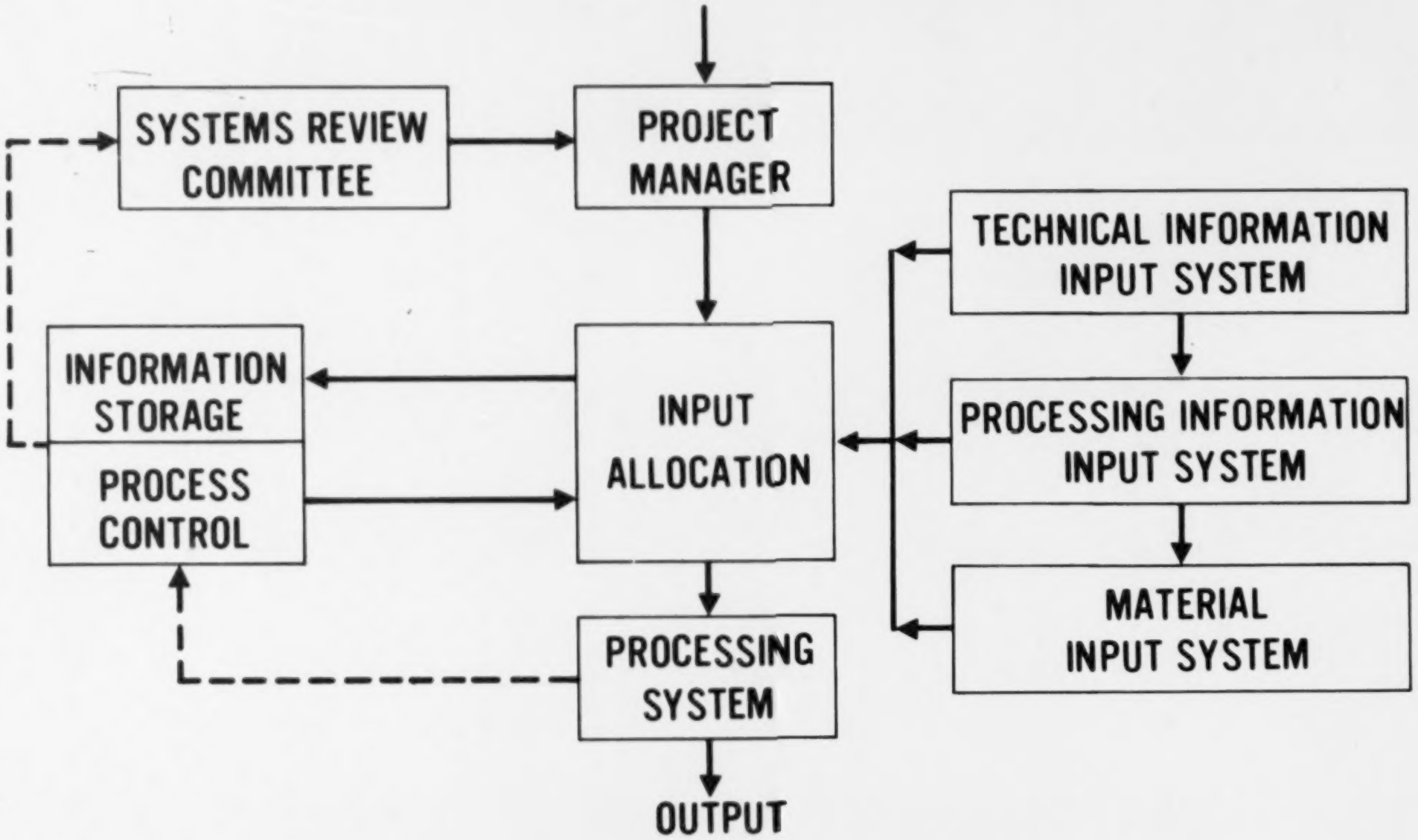
A management information system can serve the control function need to a limited degree. In addition to serving the control function for developing summary and exception reports as required, it is also feasible to envision an MIS performing in the business operating system domain: The operating system model which is a subset of the overall flow in the business organization is illustrated from a functional viewpoint in the following figure. (Reference 20.)

There are many intellectual decisions required in the business operating system as shown in this operating systems model. A major trend in MIS systems is to provide a transition from manual to automated procedures for both the control system and the operating system functions.

With this in mind, it can be seen that there is a centralized input allocation function in the business operating systems model which receives as inputs (1) technical information, (2) processing information and (3) material information. The input allocation function is, in general, coordinated with both information storage and process control modules. This type of input interface with storage and process control procedures or algorithms constitutes a basic MIS t[illegible]e or philosophy.

It is within this environment that a marketing data base can and should be constructed. The information input procedures should be a subset of overall MIS input procedures. When input is entered into the corporate MIS, no distinction should be made concerning its potential use, e.g. in a marketing or a financial data base application. The economy of scale approach which is suggested for the construction of marketing data base is that each of the source input elements be treated as a typical MIS entry. This approach enables large volumes of corporate information to be entered on a routine basis under standard procedures for audit and accuracy.

AN OPERATING-SYSTEM MODEL



Reference 20.

B. TYPICAL DATA BASE FILE SET

From a broad MIS viewpoint, a typical managerial data base file set structure is illustrated in the following figure. (Reference 21.)

The construction of a marketing data base can include elements of the entire MIS data base at given times. It is entirely feasible to have several marketing data bases defined within the corporate common data base. For example, each of the functions presented for marketing can have their own specific and inter-related data bases:

(1) Sales forecasting and projection

(2) Marketing and research

(3) Advertising

(4) Strategic marketing planning

(5) Long range planning

This type of multiple marketing data base relationship capability is entirely feasible and the list is typical in that other marketing data bases can be developed, e.g. new product development.

The trend toward specialized data bases that are conceptually defined within the table oriented software requirement indicates an ability to have marketing data bases for different executives with specialized and not necessarily identical marketing information needs.

C. TYPICAL BUSINESS TRANSACTIONS

The common data base requirement can be viewed as a multiple use of data in the time-phased cycle of business operations and in business transactions. An overview of a typical business transaction (and government, when modified), and the implied marketing data base requirement is shown in the next figure. (Reference 22.)

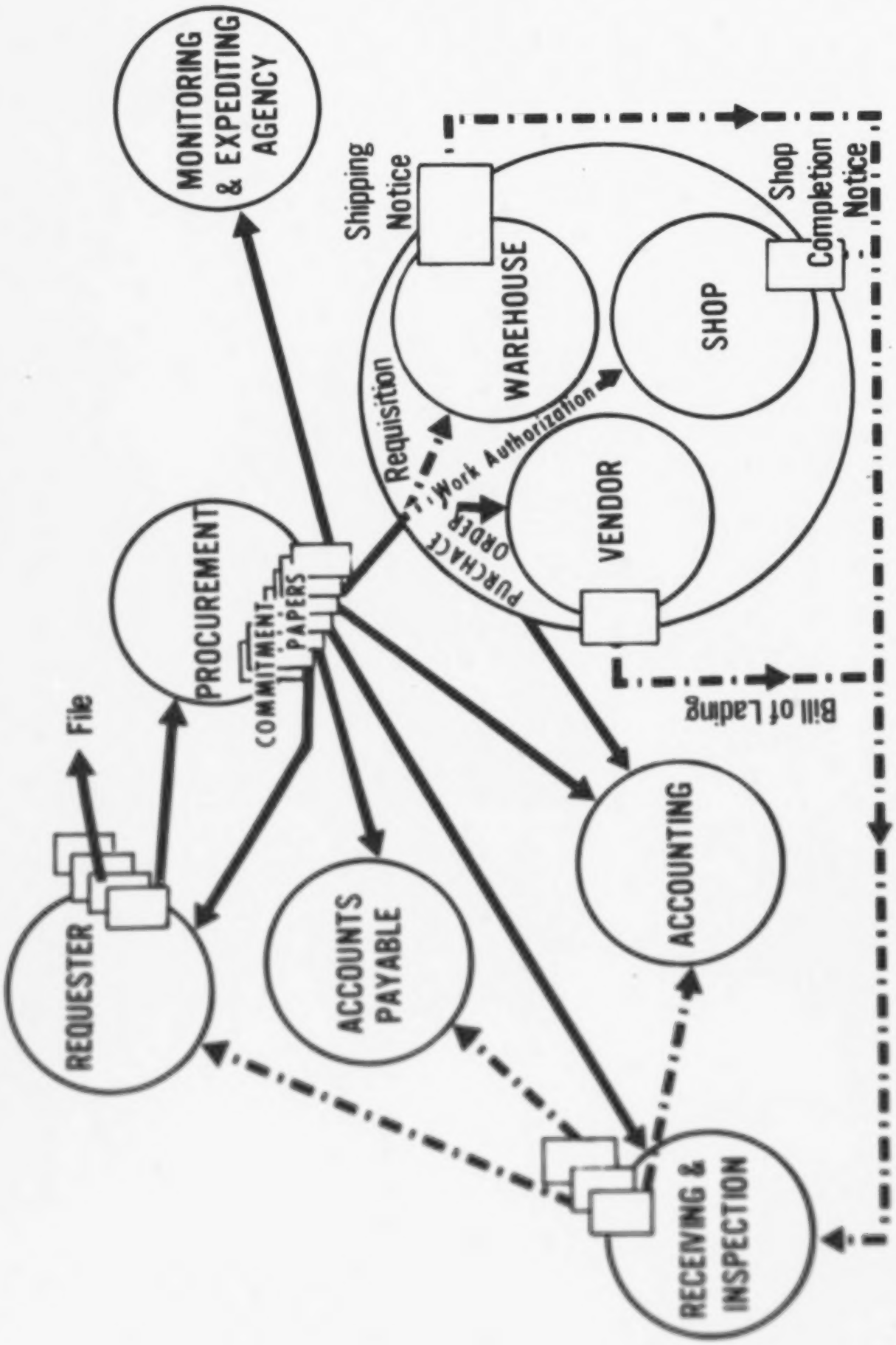
The corporate MIS is representative of a vehicle that can be utilized to simulate in the computer the paper movement for a typical business transaction. The implication here is that the marketing data base application will represent marketing paper movement inside the computer with considerable timeliness

A TYPICAL MANAGERIAL DATA BASE FILE SET STRUCTURE

- FINANCIAL DATA
- BUDGET/FORECASTING/MANAGEMENT SCIENCE
- SALES AND PROFIT
  - — PRODUCT (CURRENT, OLD, NEW)
  - — COST CENTER
  - — GEOGRAPHIC
- DISTRIBUTION
- PROCUREMENT/INVENTORY
- ACCOUNTS RECEIVABLE/PAYABLE
- PRODUCTION SCHEDULING /REPORTING
- PERSONNEL/PAYROLL
- MANAGEMENT INQUIRY PROCESSING

Reference 22.

THE MOVEMENT OF PAPERS UNDERLYING A BUSINESS TRANSACTION



improvements as well as accuracy and related benefits to the marketing management. Elements in the figure above are indicative of marketing relationships such as (1) completing the sale, (2) monitoring and expediting sale, and (3) related interfaces with the customer.

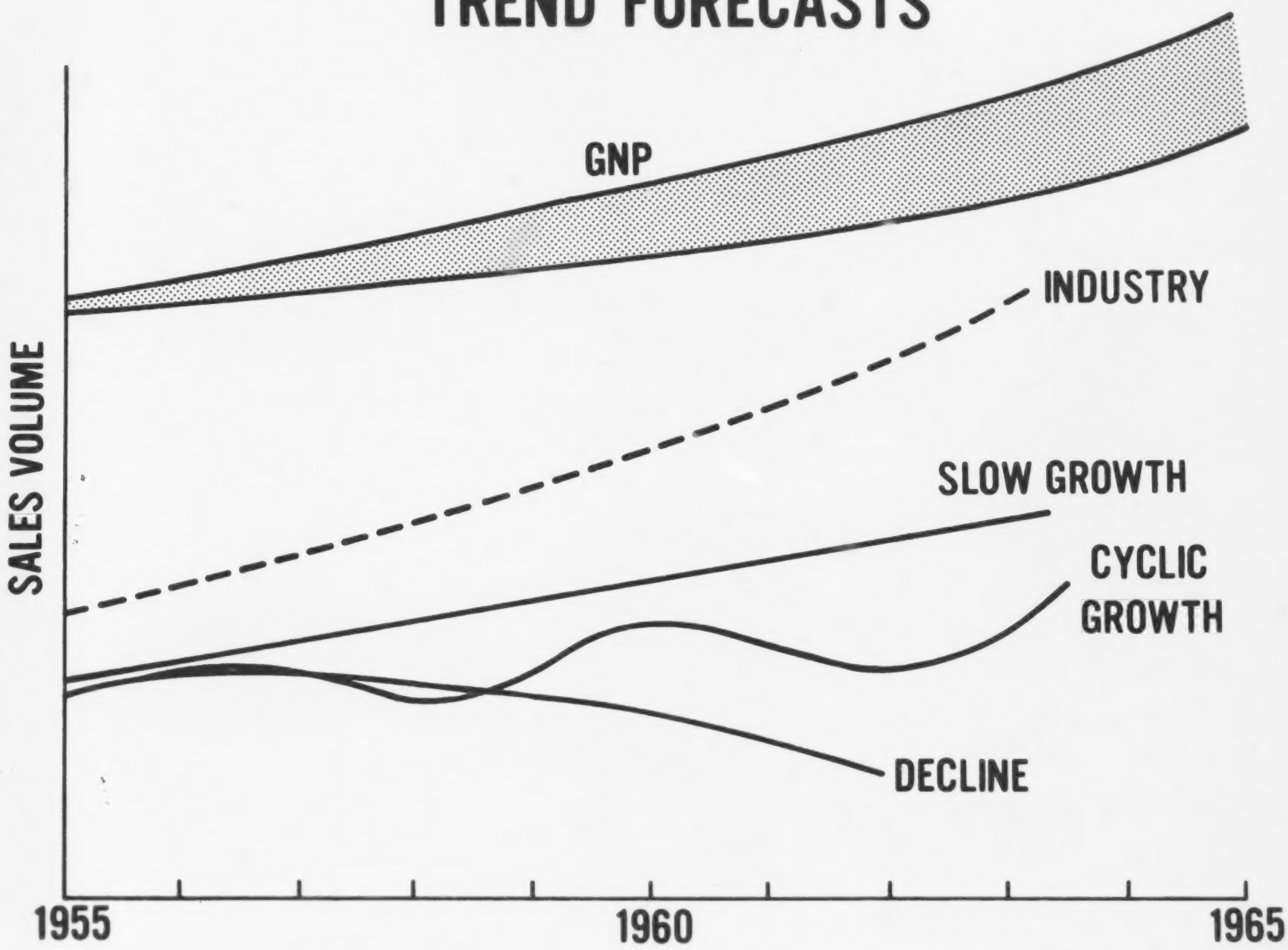
D. FORECASTING EXAMPLE

Another dimension of application criteria for use in investigating the data structure for design element is the time-phase nature of data in the MIS. By definition, an MIS should contain historical and managerial or future-directed data. In its broader sense, the future-directed data is introduced in the budget category (Reference: Carrithers and Weinwurm, BUSINESS INFORMATION AND ACCOUNTING SYSTEMS, Chapter 34).

The use of the data in terms of the appropriate time frame is also important for design consideration and particularly important for marketing data base application in such areas as sales forecasting and projection. Decisions have to be made concerning the length of time the historical information will be retained at each level of a hierarchical data base system with final retention possibly in off-line magnetic tape.

In projecting data and using budgets and beyond-budget information, consideration should be given to each data category. For example, long-range trend forecast data can be stored concerning the Gross National Product (GNP), industry, and specific company sales by product. A long-range set of possible curves is shown in the following illustration. (Reference 23.) This figure is indicative of market forecasting applications for the data stored in a marketing data base.

TREND FORECASTS



VI. MARKETING DATA BASE DIMENSIONS

A. PREDICTION AND CORRECTION

The trend forecasts curves that were introduced in the preceding figure illustrate a need for prediction when utilizing the marketing data base. The prediction is modified by actual experience and result in feedback correction.

The basic marketing management process and information flow is illustrated in the following figure. (Reference 24.)

Predictive results of a marketing program formulation strategy set are indicated by the a) feedbacks. The b) feedbacks are the correction or evalutive feedbacks that result from analysis of consumer behavior and the market environment.

This prediction and correction cycle is fundamental to the marketing data base application within an MIS.

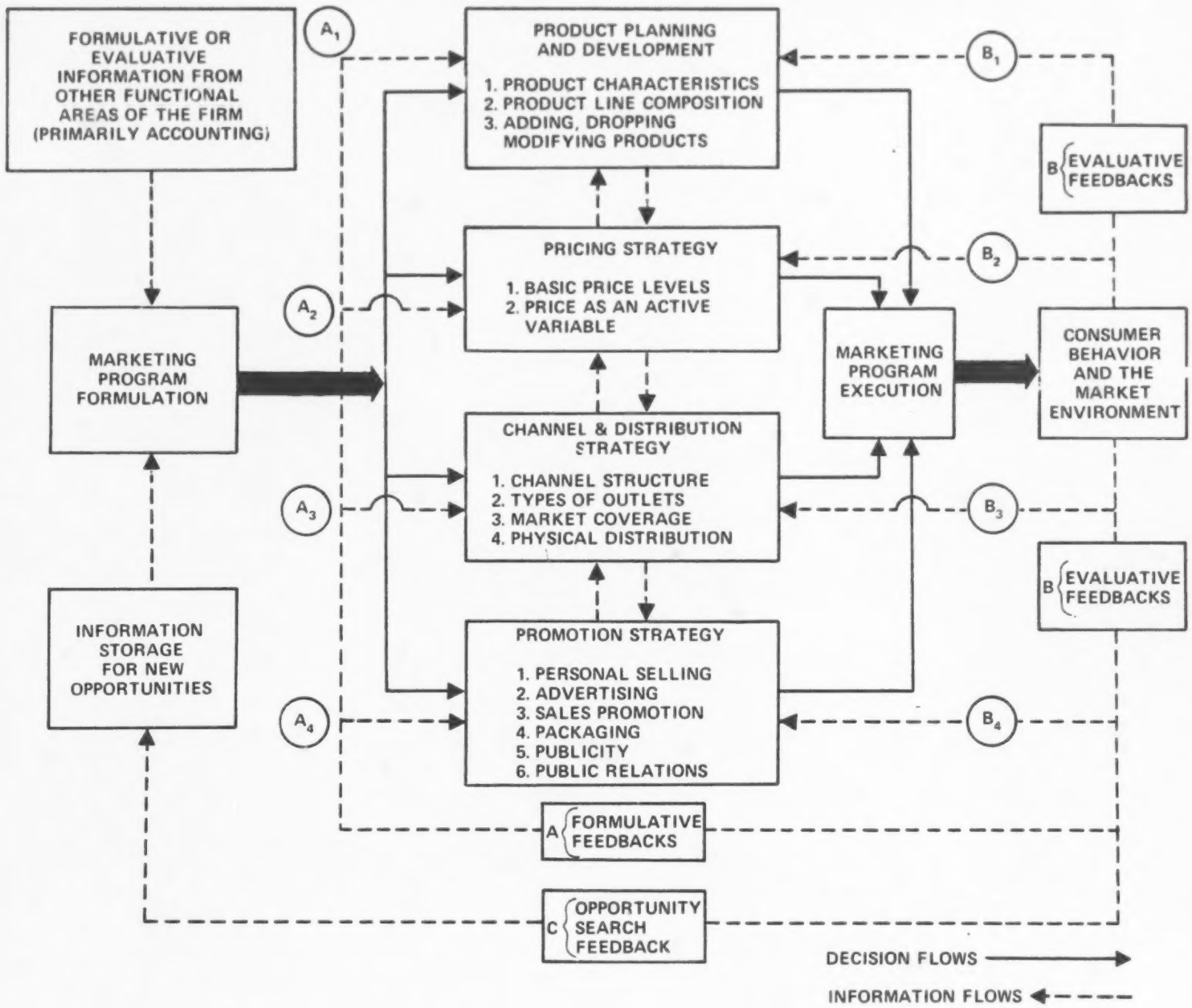
B. NEW PRODUCT DEVELOPMENT

The new product development program is illustrated in the previous figure by the opportunity search feedbacks or the c) feedbacks. These feedbacks result in the development of new products and the resulting new marketing program formulation. Marketing data base construction techniques require provision for the introduction of new and the deletion of old product data elements.

C. COMPETITIVE LEAD TIME

Effective operation feedback control mechanisms presented in the preceding figure and the basic functions relating to marketing enable the corporation to achieve competitive lead time in the introduction of new profitable products and to maintain a marketing leadership position within a given industry.

THE MARKETING MANAGEMENT PROCESS AND INFORMATION FLOW



D. DECISION MAKING

The contribution of an effective marketing data base is to economically assist management in the timely achievement of competitive decision-making processes.

Each of the marketing data base functions in the figure above are contained in MIS functions.

VII. TYPICAL DATA BASE CONSTRUCTION AND SEARCH TECHNIQUES

A. OVERVIEW

Data management systems have been introduced as consisting of the following general features (Op. Cit. Minker and Sable):

1. User Language
   a. Query Language
   b. Job Specification Language
   c. Data Description Language
   d. File Maintenance Language
2. File Structure
3. System Processing Capability
   a. Data Manipulation
      (1) Search Logic
      (2) Data and Directory Maintenance
      (3) Output Formatting
      (4) Job Management
4. Programmer Interface
   a. Procedural Language
   b. Program Micro Library
   c. System Program Library

Each data management system consists of a selection of features within the four interrelated categories above. In order to provide the reader with a highlight of typical data management system operations, this section introduces representative data base construction and search techniques.

Before the discussion is presented, an outline of the components of one data management language is provided to indicate the scope of a

typical data management system. (P.J. Dixon and Dr. J. Sable, DM-1, A Generalized Data Management System, Spring Joint Computer Conference, 1967):

A. System Languages
   1. Job Specification Language
   2. Query Language
   3. Data Definition Language
   4. Procedural Language
      (JOVIAL, COBOL, and assembly languages).

B. Design Features
   1. Modular Design
   2. Restructuring of Data

C. Functional Elements
   1. Data Base
   2. Directories (location)
   3. System Programs
   4. Supervisor Program
      a. Job Request Process
      b. System Control

DM-1 is broad in scope and as a typical large-scale data management system it requires extensive user training before it can be adequately implemented.

The balance of the discussion in this section serves as an introduction to basic concepts that can be found in data management languages. An emphasis is placed upon typical data base construction and search techniques for factual data.

Textual or nonformatted data that can be searched by Boolean techniques are not readily processed in current data management systems. Specialized software systems are required for this type of citation search function (Ref: W. T. Brandhorst and P.F. Eckert, NASA-CR-62033; N66-34085, 1966).

Only structured data Boolean search techniques are considered for the current discussion.

For review, a typical Boolean search is:

1. English Equivalent

Find all of the documents with both the term "A" and the term "B".

2. Boolean Search Equation

$Y=A \cdot B$ where Y equals the accession numbers of the appropriate documents meeting the search criteria.

B. LINKAGE

The balance of this presentation is concerned with typical MIS data base construction and search techniques which can be of value in constructing a marketing data base for structured data.

One of the terms which is vital in a data management system is the term "linkage." Minker and Sable introduce four categories of linkage which are of interest and these categories are outlined below:

1. Sequencing pointers or address identifiers (also referred to as the location of data elements) can be used to identify previous and next locations and therefore imply a sequencing linkage.

An example of a pointer concept as used in this general category of sequencing would be "chaining." In chaining, the file organization uses information that is imbedded in the data record and which contains a pointer to a location of the next record, which does not have to be stored in a contiguous manner. Chaining can be extended to the general category of branching file structures (Reference: Charles T. Meadow, The Analysis of Information Systems). Chaining, when extended into branching file structures, can in effect have several layers of related information which is linked together by pointers.

An example of a branching file structure would exist where a source record indicates that the foreign language spoken is A and the pointer A

indicates a next location which the computer goes to; and at the next location A, the code numbers for each foreign language spoken are stored; a table is utilized in the branching file structure to define the languages which are represented in the coded form. For example, the number 10 could represent English, 11 Russian, and 15 French.

2. Associative

The associative linkage concept utilizes links or pointers to develop a chain of items or data elements that have a particular value for a given information property. The associative linkage is used in a type of list and structure development (Reference: Meadow).

For example, if all of the data elements in a list refer to a type of mammal, and if each pointer contains the address of the next type of mammal, then a simple list is established. There is no necessary logical ordering in this simple list or in the more complex list structure.

A list structure is a list which contains a set of data elements or a string of data elements where at least one of the data elements is in itself a simple list. For example, a list structure for mammals might have one simple list within it that is identified by an address; this address would refer to a simple list which for example, could be dogs (a subset of mammals) and which would return to the basic list structure via a set of pointers after each of the dog categories has been identified.

3. Hierarchy

This type of linkage is used in the generic-specific relationship identifications that are found in data base vocabulary systems. A typical vocabulary system is stored in a set of dictionaries and is a part of a thesaurus. In a thesaurus, it is possible to have a multi-level hierarchal relationship as well as the broader conceptual relationships. The broader conceptual relationships can indicate relative extent of a term.

In general, these relationships are referred to as BT (broader term) and NT (narrow term). There are other types of general hierarchal relationships

such as RT (related term). Other types of relationships in the general thesaurus environment are the term use which indicates that a term is not contained as a vocabulary element for application in the data base and therefore another term must be used. For example, the term COLUMBIUM could contain a reference to use NIOBIUM.

The opposite or reciprocal of the use term is a UF (used for). In the reciprocal mode, an entry for NIOBIUM could contain a UF COLUMBIUM. These hierarchal terms are sometimes referred to as cross-reference terms for application within a thesaurus (Reference: the NASA Thesaurus, Volumes I through III).

4. Adjoining

The adjoining linkage can be an address or an implied address which establishes a relationship between different parts of a data record. The implication is that both data elements sets, namely, the basic data or source data and the related or target data items are on the same logical level. A common example of adjoining linkage is the establishment of a relationship between the fixed and variable data fields for a logical record.

These basic terms and concepts are expanded as applicable in the following discussions concerning the development of a common data base for MIS and associated constituent data bases, e.g., marketing data base.

C. DICTIONARY

Dictionary development and thesaurus creation is a vital part of the data-base vocabulary control and translation process. A dictionary is illustrated in the following figure. (Reference 25.)

The dictionary presents a-word-to-code dictionary which is in effect a definition table that indicates a proper code usage. For example, "aircraft" or "airplane" use the internal code 01.

More profound dictionaries that can process concepts (e.g. a group of synonyms that define homograph meanings) are usually not found in current data management systems. One such dictionary that is under development for a man-machine interactive query language is illustrated below (Ref: Dr. David Fox, Leasco Systems & Research Corporation): (Reference 25a.)

A WORD-TO-CODE DICTIONARY

WORD	CODE
AIRCRAFT	01
AIRPLANE	01
CALCULATOR	02
CANINE	03
COMPUTER	02
DOG	03
PLANE	01

WORD-TO-CLUSTER DICTIONARY

CONCEPT WORD	EQUIVALENT CONCEPT WORDS	CODE
MEMORIES	AUTOMATIC MEMORIES, CORE, INTERNAL, INTERNAL STORAGE MASS MEMORIES, MASS MEMORY, MEMORIES, MEMORY, MEMORY CORE, MEMORY SYSTEM, STORAGE	80
MEMORIES	CAPACITIVE MEMORIES, MEMORIES, MEMORY, READ ONLY CAPACITIVE MEMORIES	87
MEMORIES	INDUCTIVE, INDUCTIVE MEMORIES, MEMORIES, MEMORY	88
MEMORIES	IBM PHOTOSCOPIC DISC, MEMORIES, MEMORY PHOTOGRAPHIC, PHOTOGRAPHIC MEMORIES	89
MEMORIES	HARDWARE, HARDWARE ORGANIZATION, MEMORY, MEMORY ORGANIZATION	107
MEMORIES	CARD MEMORY, MEMORIES, MEMORY PERFORATED CARD MEMORY	224

The first dictionary presented a typical structured definition table. In more profound literature thesaurus development and concept search applications, more complex dictionaries can be useful for the user.

In the example presented above, the homograph "memories" is precisely defined in terms of a given set of user applications. This word-to-cluster (e.g. group of synonyms) is a variation of a word-to-code dictionary. For example, the word "memories" codes to the number "80" for a given search application.

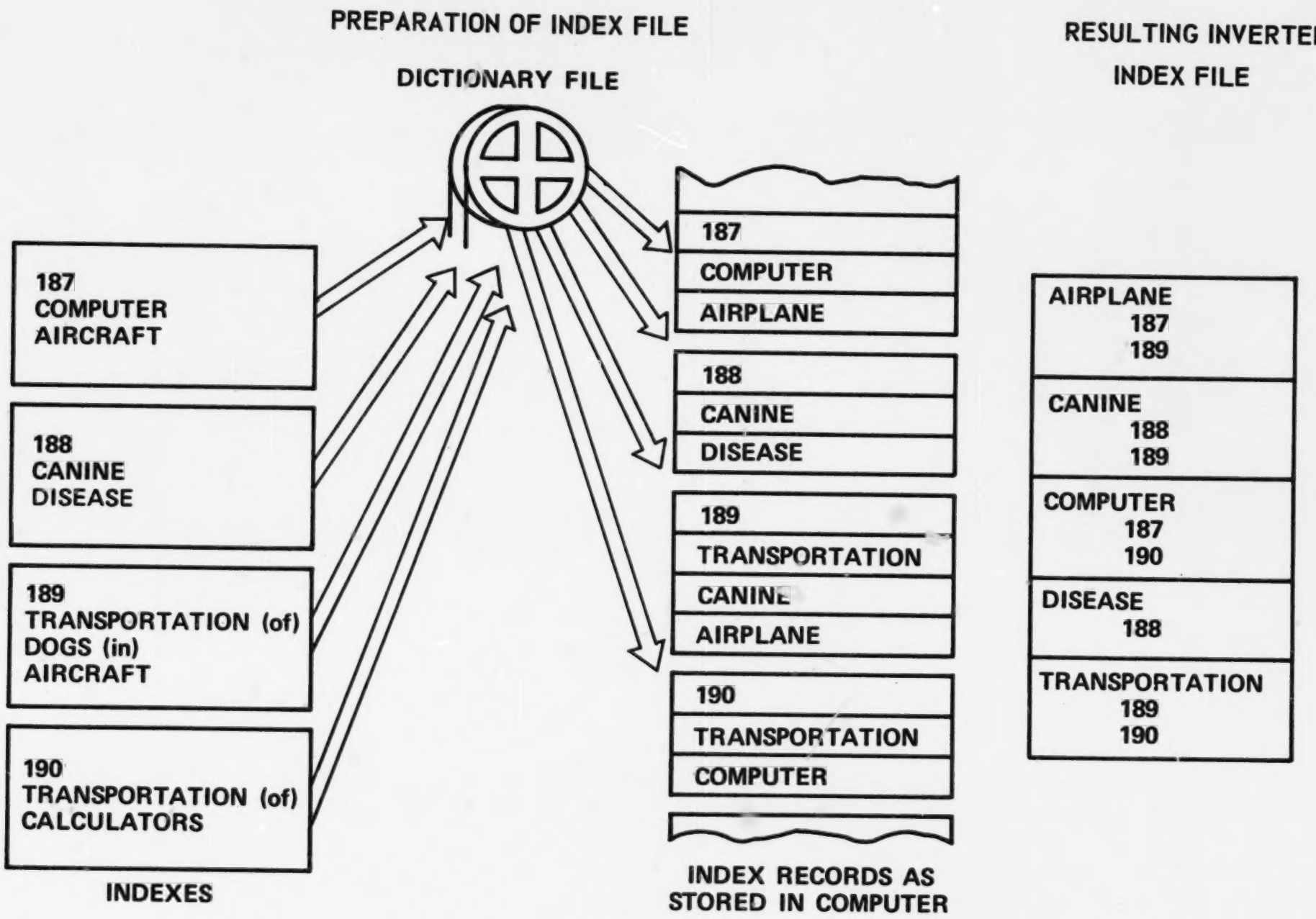
D. INVERTED INDEX FILE

A figure is presented below concerning basic data base organization techniques to introduce the inverted file structural concepts which are used in coordinate indexing and search techniques for structured data. (Reference 26.)

The inverted file index is prepared from a standard index file as shown in the preceding figure. The standard index file is often called a serial index file, and is very similar to development of card catalogs in the library system. In the index file (serial) development, the documents such as Nos. 187 and 189 are entered in translated format in a dictionary file and output as serial records with appropriate word-to-word dictionary equivalents. This example is shown with a magnetic tape implication as are most of the basic examples in data base organization techniques. Sight should not be lost of the fact that the dictionary file and related tables will normally be found on extended core and secondary rotating storage devices (for example, drums, disks). The trend is away from the basic or traditional magnetic tape orientation. In particular, inverted file utilization is more applicable for direct access methods because of a rapid search capability.

The power of the inverted index file is in the fact that it can be used with coordinate, or Boolean search equations to focus rapidly and precisely upon the potential candidate documents without performing an expensive serial search of the entire data file. The example of the inverted data file which was prepared from the sample index or serial file in the above figure is inverted in the sense that the terms are posted as separate records and all documents

INVERTED INDEX FILE



Reference 26.

that contain those terms are posted against the terms. For example, "airplane" is found on two documents; therefore, two documents are posted against the term "airplane," i.e., 187 and 189.

E. FILE MAINTENANCE

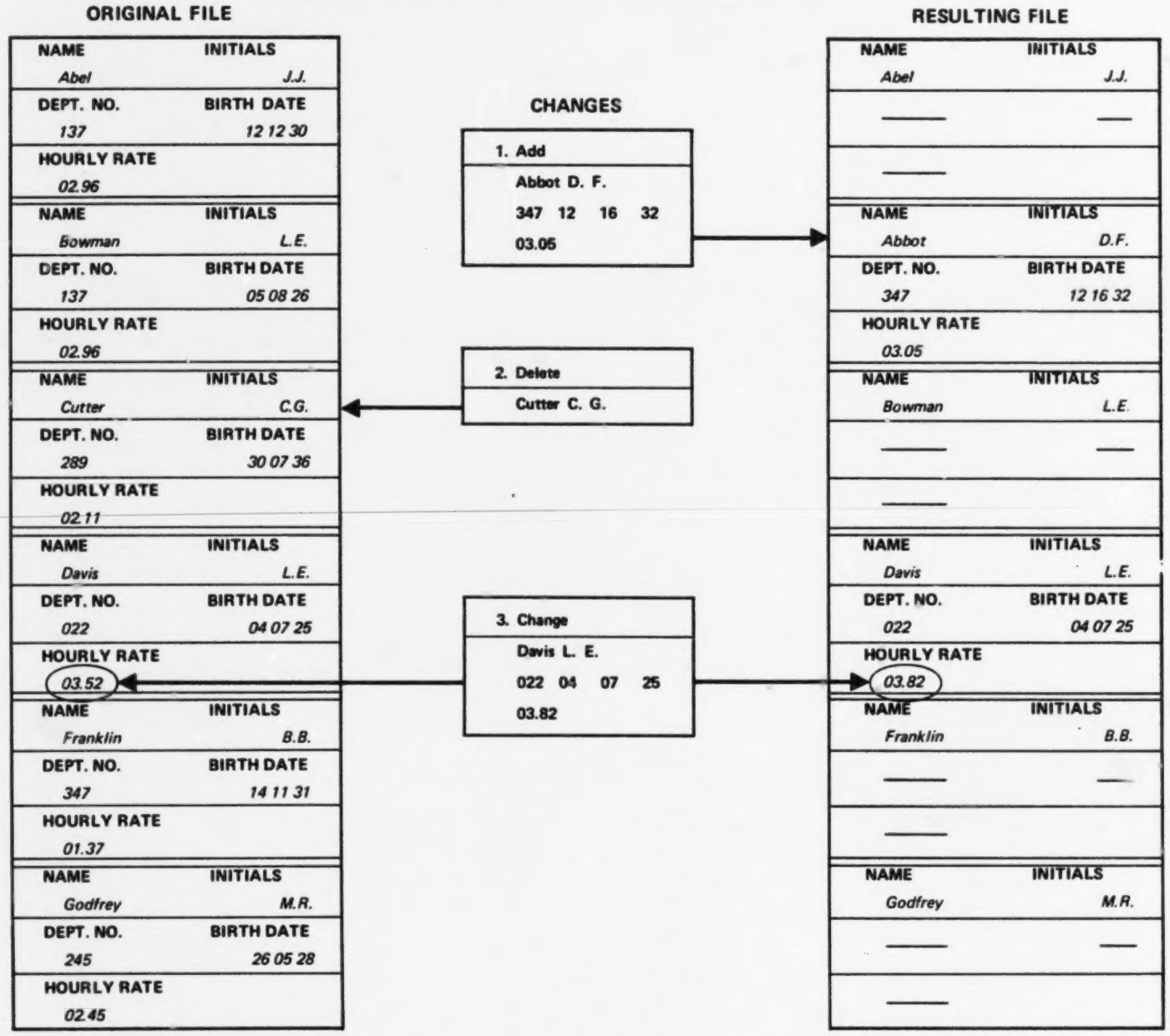
Two figures follow which introduce file maintenance and interface with file searching. (References 27. and 28.)

File maintenance transactions as illustrated in the first figure show three basic functions: (1) add, (2) delete, and (3) change. There are many other possible transactions such as (4)transfer. The (1) add in this example is adding the name "Abbot" with the initials"D.F." and appropriate numerical information. The (2) delete indicates a deletion of similar information for a man named C.G. Cutter. And the (3) change is for an hourly rate for a man named L. E. Davis. In each case, the file maintenance or update procedure is shown as an original and resulting file to give a review of the process. In the case of the delete there is no resulting file record, i.e., a null file record.

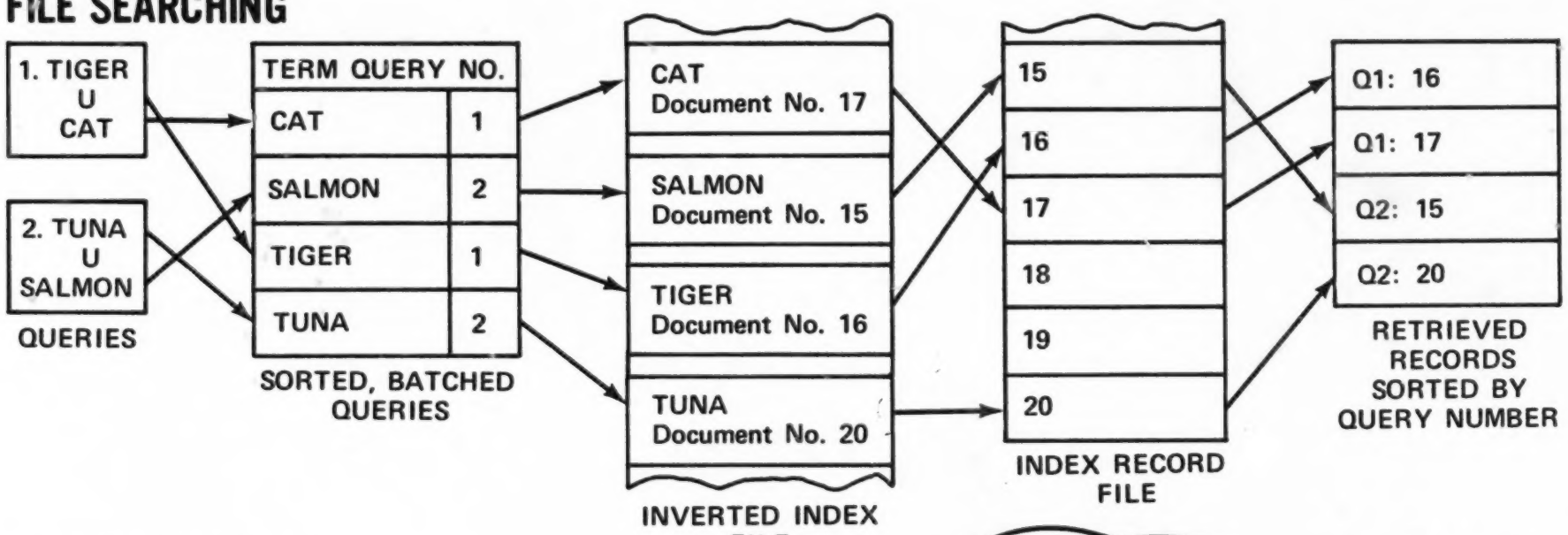
File maintenance as related to file searching is illustrated in the second of the two figures. The bottom part of the figure illustrates a file maintenance transaction which is fundamental in large-scale data base systems. The significant point of this file maintenance is that new index records (for example, those that contain the key words "Pine," "Fur," "Zinnia" and "Aster" are entered into both serial or linear index files and inverted files. In developing a common data base, it is necessary in large-scale systems to use both index or serial and inverted files and combinations thereof. The file maintenance requirement is for a comprehensive algorithmic capability to coordinate every new index record with its appropriate format and position in each applicable index and serial record and inverted record. There should be a one-to-one relationship for every term that is in an index record, i.e., there should be an equivalent term in the inverted index file.

A multiple mapping of record numbers can occur from the index record file to the inverted file since each index or record number is replicated according to the number of terms an index record contains. For example, in the

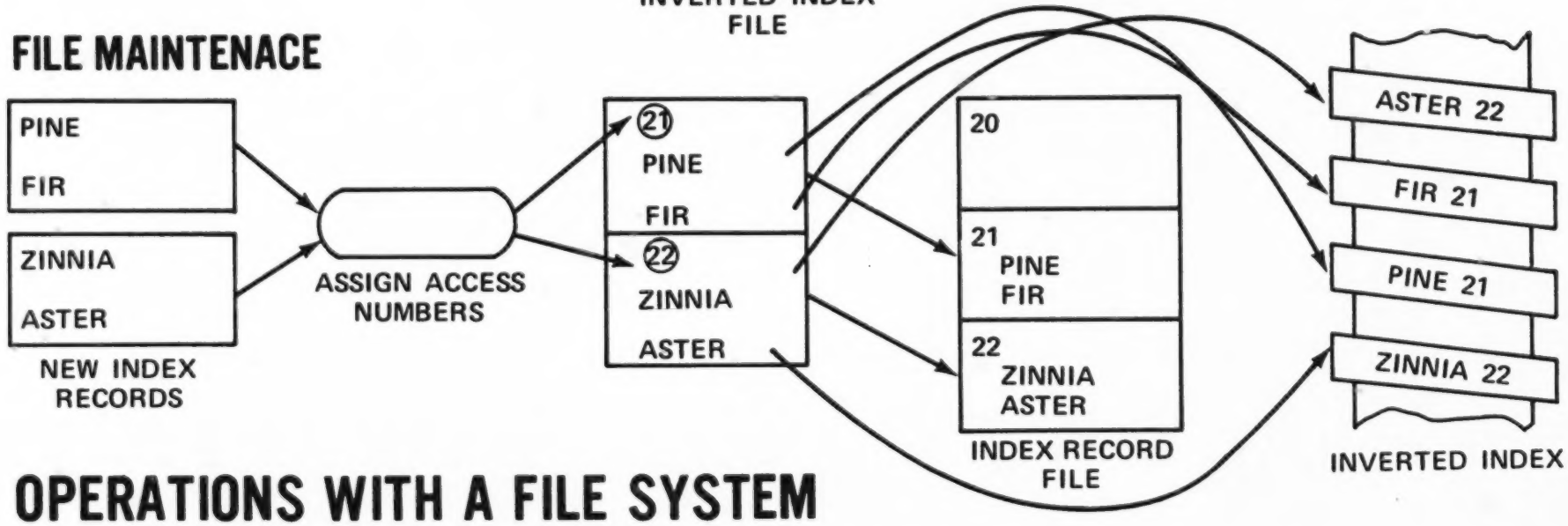
FILE MAINTENANCE TRANSACTIONS



FILE SEARCHING



FILE MAINTENACE



OPERATIONS WITH A FILE SYSTEM

Reference 28.

document number or index No. 21, one contains the key words "Fur" and "Pine", so 21 is replicated twice, once under "Fur" and once under "Pine." It is important for the computer system to maintain accurate records during file maintenance so that all of the index record files are mapped correctly into the inverted index files. The implication for other file maintenance transactions applies, such as deletion, modification and so forth. All file maintenance operations which involve one or more records have to be modified on each applicable record.

The search above using a combination of the inverted and indexed records shown at the top consists of two search inquiries. The symbol "U" represents a plus or Boolean union of two terms. This is commonly referred to as a logical "or." An upside-down "U" represents an intersection of two terms and is usually represented by a dot and refers to a logical "and" operation.

F. FILE SEARCHING

1. Inverted File

There are a variety of search techniques available for use with common data base systems. Many of the search techniques have been developed from traditional coordinate indexing search systems such as those at the NASA Scientific and Technical Information Facility (Reference: A Guide by W. T. Brandhorst and Philip F. Eckert, June 1966: NASA CR - 62033). There are many new search techniques which are being developed utilizing these basic equation search methods where the search strategy is established as an equation.

An example of an equation search using the NASA literature methods is presented below in an English format:

"Find all the documents that relate to the testing of material but avoid its destruction."

This English equation is further quantified in the semi-structured form:

"Find all those records that contain the key words or terms 'non-destructive' or ultrasonic' and 'structure' or 'structural' or 'material' or 'honeycomb' or 'forging' or 'vessel' and 'testing' or 'inspection'."

The strategy involved in this type of coordinate search using a Boolean equation equivalent to the English statement above would be to define a logical intersection of three elements: (1) kinds of testing that do not destroy the tested item, (2) the kinds of items to be tested, and (3) the act of testing. This type of search procedure is provided in a document titled Illustrative Literature Searches Using NASA Data, Leasco Systems & Research Corporation, 1968).

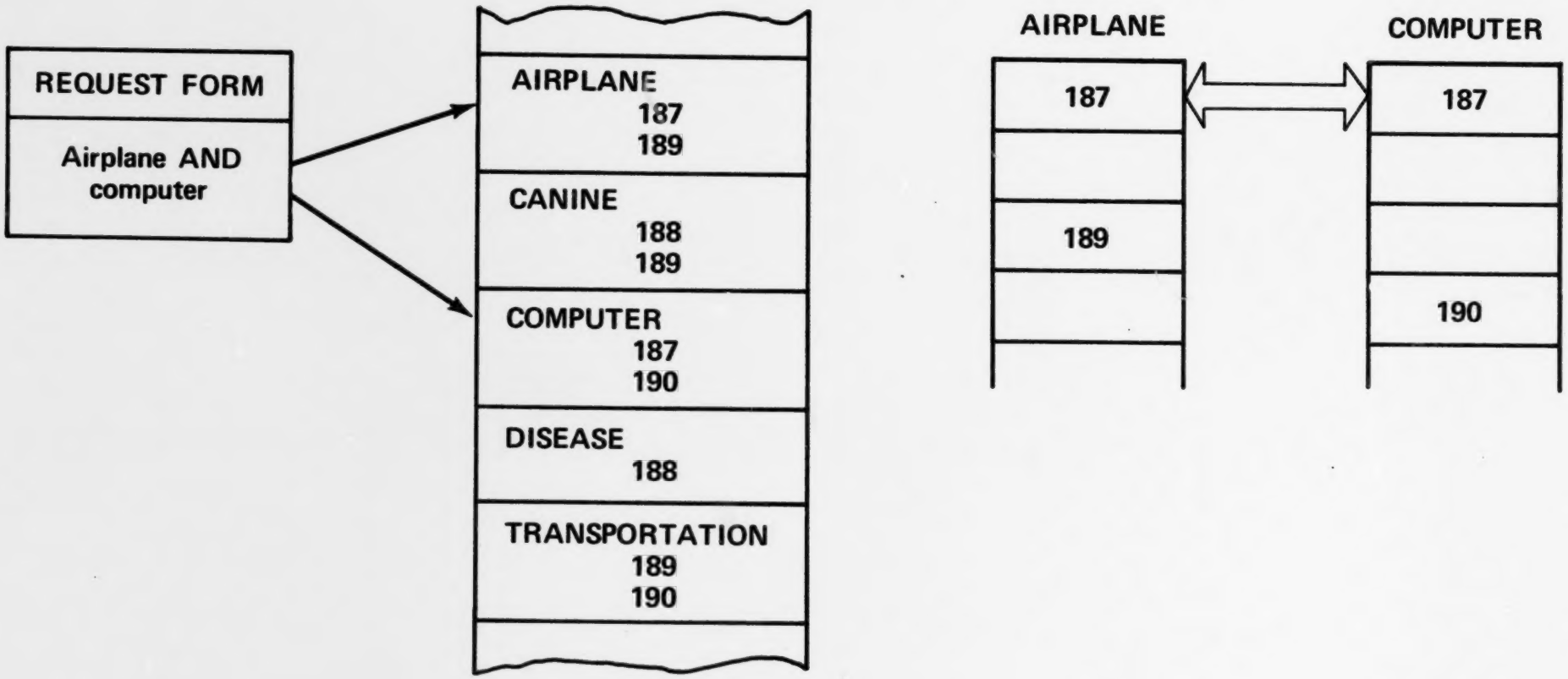
Equation searches can provide a variety of complex logic that is translated into relatively powerful search algorithms tailored to individual inquiries. There are Boolean search software systems developed which can be modified for on-line information systems to operate in a direct access mode. Rather than present the more complex search approaches, an emphasis in this discussion will be upon fundamentals of search techniques which can be used in common data base environment.

2. Direct Access

The first example is that of simple search using an inverted index file and is presented in the figure below. (Reference 29.)

In this sample illustration, a request for all those records containing both the term "airplane" and "computer" (translation through a dictionary process implied) is illustrated. Document No. 187 meets this requirement.

USE OF AN INVERTED INDEX FILE



Reference 29.

With this type of inverted file search procedure in mind, the direct access searching techniques are introduced by a simple example in the next figure. (Reference 30.)

In the case above, the request is for all those documents that contain both computer and airplane. A strategy is illustrated as a possible approach. In this case, the direct access strategy consists of searching the inverted file for only one of the terms connected by an "and," namely, the term "airplane." After the inverted file is searched, the documents or hits which are represented by the term "airplane" are then further searched for the second term which is connected by the logical "and" namely, "computer." The retrieved records that can be displayed are those records which meet the search equation requirement for "computer" and "airplane." In this case, record No. 187 meets this requirement.

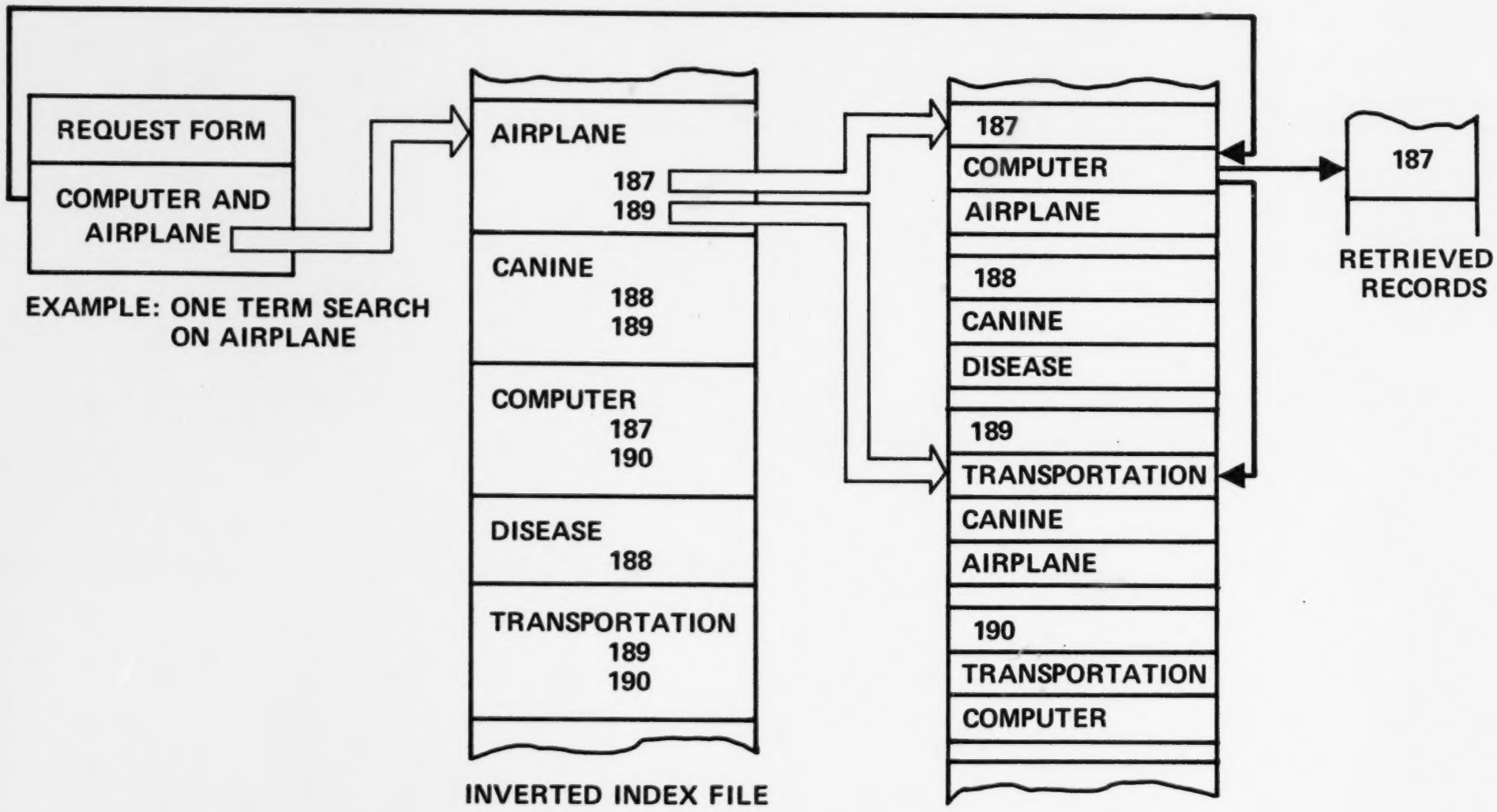
The example which has just been presented for direct access searching illustrates some of the fundamental relationships which are possible between inverted and indexed files and index and serial files. At one time in the development of information retrieval methodology, there was a school of thought which tended to judge the advantage of index or serial files versus inverted index files. This school of thought is gradually giving way to a more complex approach as is implied in the example above. The interrelationships that can be used for effective organization of file sets in a large-scale common data base are utilized in a query language so that a wide variety of searches can be achieved. These searches utilize the advantages of both inverted and serial file construction methods.

The discussion in this section concerning search is somewhat limited in terms of fundamentals. On-line MIS systems have many more considerations which should be evaluated (Reference: A TRANSITION TO ON-LINE COMPUTER, by Fred Gruenberger.

3. Combined

Three basic types of file search techniques are presented (Concluding references from Charles T. Meadow) to demonstrate a variety of techniques

DIRECT ACCESS SEARCHING



Reference 30.

which are available for consideration in utilizing and designing an MIS software system. The first technique is presented in the following figure.(Ref.31.)

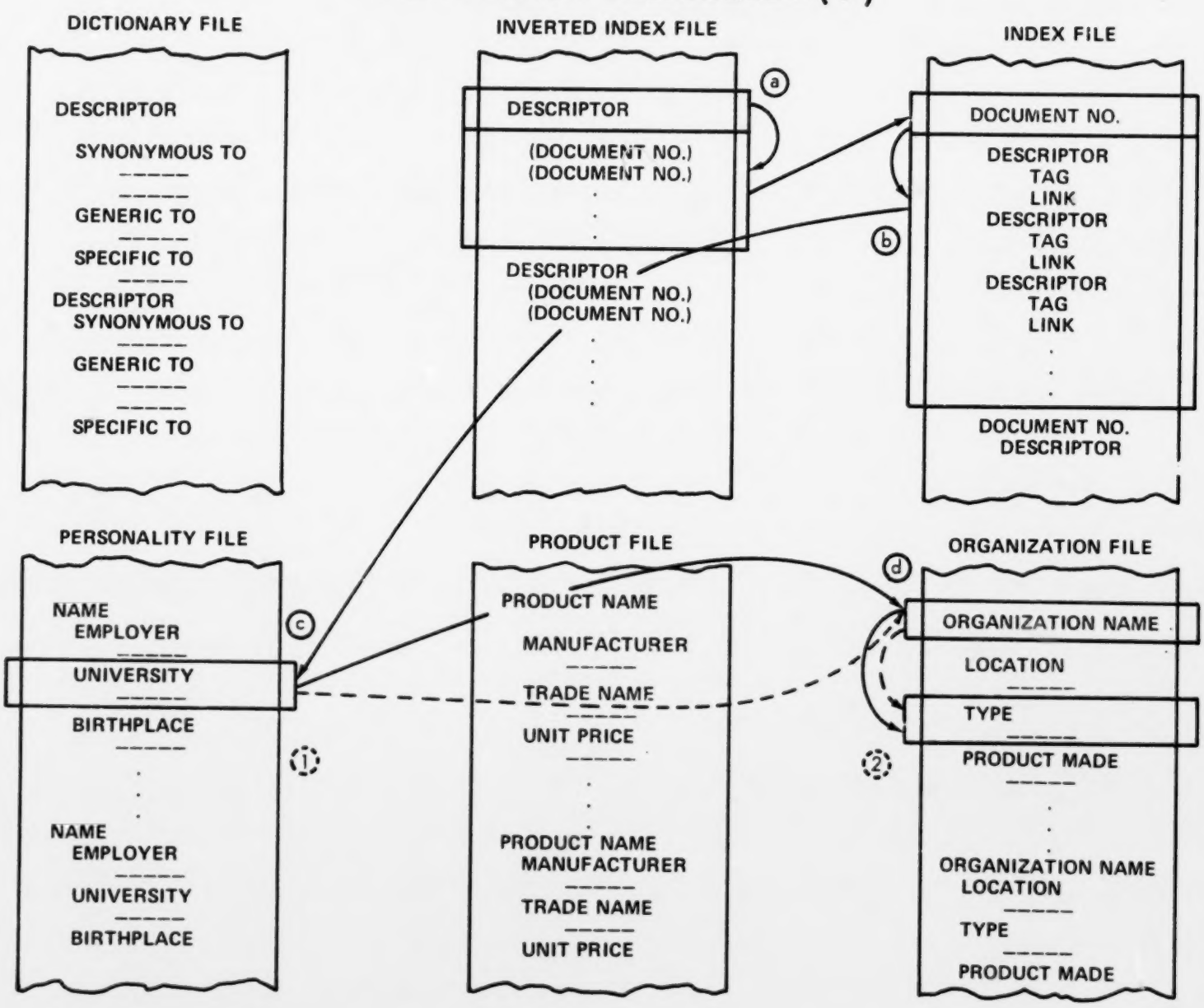
This figure which is titled "File Search Techniques I" illustrates a search utilizing essentially several different types of files in the inverted and serial classification categories. The logic through the searches varies in each of the three examples used to conclude this search discussion. However, the structure of the files is assumed to remain the same and constitutes a constant for these three examples.

The first question which is resolved in file search techniques (I) is that of finding a list of all Harvard men who are in the electronics industry. There are possible Boolean or equation methods; in general, the examples in this section approach the precision of fact retrieval as contrasted with document retrieval for solving this search question in a well-structured literature file. In the example presented, there is no file sequence on university or industry type and there is no highly structured literature file which would be readily available for a coordinate search system per se.

The general approach followed in the file search technique (I) is to search the inverted index file in step a to find all of the index records that do contain "Harvard." These document numbers are then searched in step b (the documents are in the search or index file). In step c, the personality file is searched for all those names of Harvard-associated people that were derived from the index file. Upon verification that a name in the personality file meets the criteria of a Harvard man, i.e., a man who attended Harvard University, then the logic search is directed toward the organization file to determine the type of organization where the individual is employed and evaluate whether or not he is in the electronics industry.

The final result is a list of all Harvard men who are in the electronics industry as represented by these files. This approach which is illustrated above is perhaps more precise than might be the standard coordinate search utilizing an inverted file base as the main search domain. The method utilized in this example is faster than a serial search (as indicated by the dotted #1,

FILE SEARCH TECHNIQUES (I)



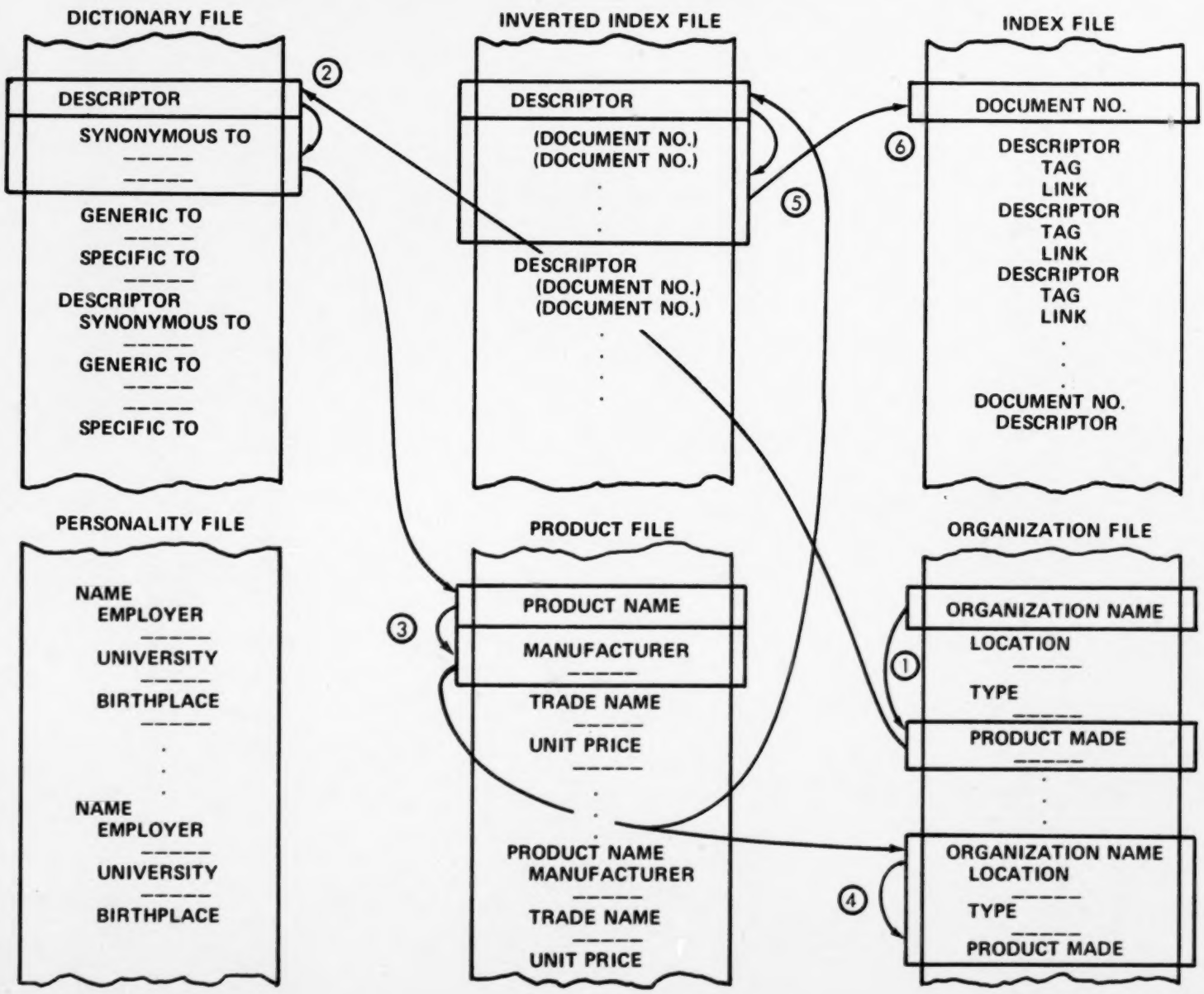
#2 in the figure); or from the organization files which would require at least a full sequential search of the personality file.

Utilizing the same six files to constitute a set of files for this sample search, a second question is posed. The question is, "Who competes with ABC Company, and what information do we have concerning their operations?" In this sample search, six steps are illustrated as a possible approach in the following figure.

The first step is (1) to search the organization file for the ABC Company name and to retrieve a list of the products made by ABC. In a more complex search system, it would possibly be necessary to determine which are the major products and continue the search only on the major products. The second step (2) is to take the list of products derived from the organization file and match them against the dictionary or vocabulary control file for the system and to find all appropriate terms and synonyms. The third step consists of (3) searching the product file with the terms derived from the dictionary or thesaurus file to retrieve the names of each organization that produces products in this product area in which ABC is active. Once again, there is an over-simplification, and screening would be performed most likely in a manual method to determine relative strength of competitors. If this type of search is part of a high-volume operation, then it is possible to be coordinated with searches of financial data base reports concerning companies in the file so that certain pre-screening can be done; for example, of a dollar value of assets, net earnings, or net sales. (Reference 32.)

Now that the competitors are identified, the computer system would turn toward the (4) organization file once again to retrieve a summary of organization and related geographic information on each competitor. Steps (5) and (6) indicate a more in-depth conclusion to the search: The introduction of the corporate name of each competitor into each of the inverted files and generation of copies of all documents in the index file that pertain to each competitor.

FILE SEARCH TECHNIQUES (II)



The search strategy utilized in the second example presents a method of utilizing different types of files that are sorted on different sort keys or fields and entering these files on a field that is their sort key.

The third or last search utilizing the six basic data files is titled "File Search Techniques (III)," and is once again a basic search technique which is used to demonstrate some strategies which are of possible interest in utilizing files that are sorted on different keys. The question in the final example in search techniques or the search phase is, "What materials are available near each of five sites that are being considered for a new location within the needs of a given manufacturing plant?" The needs of this manufacturing plant consists of ten different raw materials. In the example used by Meadow, nearness is considered the same state and reflects the simplification theme which has been used in each of these three examples. This search is illustrated in the last figure.

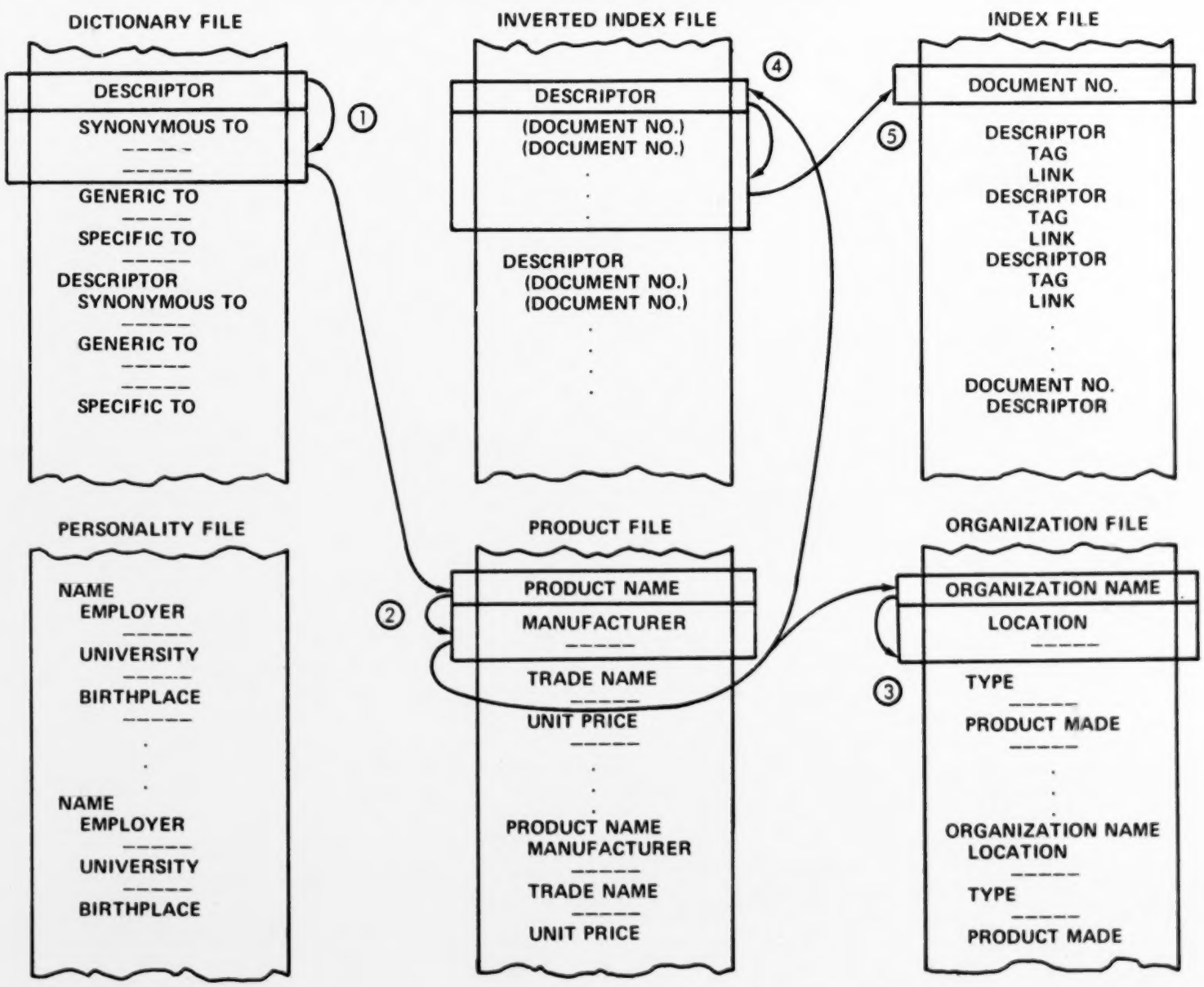
The five steps in the example consists of a sequence of searches similar to the preceding. The first step (1) is the entry of the dictionary or thesaurus to determine synonyms and appropriate names for each of the needed products.

The last example in file search techniques is presented in the following figure. (Reference 33.)

When the product file (2) has been searched, the names of all the appropriate manufacturers will be generated. In step (3), the manufacturers are searched in the organization file to determine location, which in this case would be state-oriented. Steps (4) and (5) can be utilized to secure additional information concerning each manufacturer by use of the inverted serial index file relationships discussed previously.

The search analysis which has been presented so far is in the general area of developing a common data base for a management information system. There has been a certain file replication within each of the last three examples for the six files used as an illustrative set of files. The illustrative set of files or volume consisted of (1) dictionary file, (2) inverted index file, (3) index

FILE SEARCH TECHNIQUES (III)



(or serial) file, (4) personality file, (5) product file, and (6) organization file. The replication and coordinate file maintenance systems requirements for maintaining information in each of the six files (in a broader case) place a considerable burden on editing, updating, and general file maintenance operations in the data management system operation.

G. SUMMARY

A basic introduction for the foundation of constructing a marketing data base has been presented. The final discussion concerned typical specific computer software techniques as applicable to the common data base within an MIS environment.

The theme of this presentation has been the integration of the marketing data base into the overall corporate decision-making process and supporting MIS common data base. Marketing data base construction procedures are therefore a subset of the corporate common data base construction procedures.

Data management software concepts have been introduced as economical means of achieving marketing data base file maintenance and processing operations. Table - oriented software techniques were presented as applicable to representative MIS data bases such as the marketing data base.

VIII. FIGURE REFERENCES

Number	Title	Page
1.	Data is Contained in Information AMA Presentation, Harold J. Podell	4
2.	General Form of the Logistic Curve Little Science, Big Science, Derek J. De Solla Price, Columbia University Press, New York & London, 1963	5
3.	Sample Set Format Internal Leasco Document, by Barbara Ambrose.	9
*	Inverted File Example Ibid. 3.	--
5.	Data Definition Tables Ibid. 4.	10
6.	Data Definition Tables Ibid. 5.	12
*	Sample Data Set Definition Ibid. 6.	--
8.	System Tag Tables Ibid. 7.	14
9.	A Partial Example of Data Bank Construction and Operation AMA Presentation, Harold J. Podell	16
10.	Managerial Data Base Conceptual/Structurial Relationships AMA Presentation, N.W. Nilson	18
11.	Managerial Data Base Typical Field Characteristics Ibid. 10.	19
12.	The Data Base Concept Op. Cit. 9.	21
13.	Overall View of Data Management System Processing Internal Leasco Document, by G. Woolley.	22

Number	Title	Page
14.	The Data Base Concept AMA Presentation, W.K. Collins	24
15.	Elements of an Information Center AMA Presentation, Harold J. Podell	25
16.	An Automated Information Center Ibid. 15.	28
17.	An Advanced Digital Computer Time-Sharing System Op. Cit. 9.	30
18.	Data Base Management Typical Language Functions Op. Cit. 14, Collins.	34
18a.	Structure of Purchasing Data in DM-1: Example Spring Joint Computer Conference 1967, P.J. Dixon and Dr. J. Sable	35
*	Information Flow In A Business Organization The Theory and Management of Systems by Richard Johnson Freemont Kast & James Rosenzweig, McGraw Hill, N.Y., 1967	--
20.	An Operating-System Model Ibid. 19.	40
21.	A Typical Managerial Data Base File Set Structure AMA Presentation, Delanie Patrick, Jr.	42
22.	The Movement of Papers Underlying a Business Transaction The Financial Manager by Jerome B. Cohen & Sidney M. Robbins, Harper & Row, Publishers (New York) 1966	43
23.	Trend Forcasts Long-Range Planning For Management David W. Ewing, Editor Harper & Row, Publishers (New York) 1958	46
24.	The Marketing Management Process and Information Flow Marketing Information Systems: A New Dimension for Marketing Research, by Richard H. Brien and James E. Stafford, In THE JOURNAL OF MARKETING, July 1968, Vol. 32, No. 3.	48

Number	Title	Page
25.	A Word-To-Code Dictionary The Analysis of Information Systems, A Programmer's Introduction to Information Retrieval, Charles T. Meadow, John Wiley & Sons. Inc., New York	55
25a.	Word-To-Cluster Dictionary Internal Leasco Document, by Dr. David Fox.	56
26.	Inverted Index File Op. Cit. 25, Meadow.	58
27.	File Maintenance Transactions Ibid. 26, Meadow.	60
28.	Operations with a File System Ibid. 27, Meadow.	61
29.	Use of an Inverted Index File Ibid. 28, Meadow.	64
30.	Direct Access Searching Ibid. 29, Meadow.	66
31.	File Search Techniques (I) Ibid. 30, Meadow.	68
32.	File Search Techniques (II) Ibid. 31, Meadow.	70
33.	File Search Techniques (III) Ibid. 32, Meadow.	72

* Figures will be in the next issue.

IX. SOURCE LIST

Ambrose, Barbara. ADVANCED INFORMATION SYSTEM; DATA BASE CONSTRUCTION PROCEDURES. Bethesda, Md, Leasco Systems & Research Corp. (formerly Documentation Incorporated). Limited Note LN-AIS/311/1-0, Oct. 30, 1967.

Brandhorst, W.T. and P.F. Eckert. GUIDE TO THE PROCESSING, STORAGE AND RETRIEVAL OF BIBLIOGRAPHIC INFORMATION AT THE NASA SCIENTIFIC AND TECHNICAL INFORMATION FACILITY. June 1966. NASA-CR-62033 N66-34085.

Brien, Richard H. and James E. Stafford. Marketing Information Systems: A New Dimension for Marketing Research. In THE JOURNAL OF MARKETING, July 1968, Vol. 32, No. 3.

Campagna, Joseph F., American University, Washington, D.C. (Private communication).

Carrithers, W.M. and E.H. Weinwurm BUSINESS INFORMATION AND ACCOUNTING SYSTEMS. Columbus, Ohio, Charles E. Merril Books, 1967.

Collins, W.K. Languages for Data Base Management Information Systems. (Informal presentation at AMA Seminar No. 6207-2, July 9, 1968).

Cuadra, Carlos A., Editor. ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY, Vol. 2. New York, Interscience (Wiley), 1967.

Dixon, P.J. and Dr. J. Sable DM-1-A Generalized Data Management System. Spring Joint Computer Conference, 1967.

Fox, David, Leasco Systems & Research Corp. (Private communication).

Gruenberger, Fred, Editor. TRANSITION TO ON-LINE COMPUTING; PROBLEMS AND SOLUTIONS. Washington, D.C., Thompson Book Company, 1967.

Irwin, Manley R. Government Policy Implications in Data Management. In DATAMATION, June 1968, p. 37.

Kessler, Myer M. Fundamentals of Information Retrieval Systems. (Informal presentation at AMA course No. 6533-01 on information. New York, April 29-May 3, 1968).

Meadow, Charles T. THE ANALYSIS OF INFORMATION SYSTEMS. New York, John Wiley & Sons, 1967.

Minker, J. and J. Sable. File Organization and Data Management. In Cuadra, op. cit., pp. 123-160.

Nilson, N.W. The Logical Data Base--Its Concepts, Development and Use. (Informal presentation at AMA Seminar No. 6207-2, July 9, 1968).

Office of Education, H.E.W. OFFICE OF EDUCATION REPORT GENERATOR SYSTEM; USER'S MANUAL. Prepared by Fox Computer Services, Inc. (now Leasco Systems & Research Corp., Bethesda, Md.), Dec. 1967.

Patrick, Delanie, Jr. Conceiving the Data Base: One Company's Approach to Creation and Structure. (Informal presentation at AMA Seminar No. 6207-2, July 8, 1968).

Podell, Harold J. Developing the Common Data Base for Management Information Systems: I Environment for Change; IV Developing a Common Data Base for MIS, AMA Seminar No. 6207-02, July 8-10, 1968).

Stout, Harold, Auerbach Corporation, Philadelphia, Pa. (Private communication).

Sunden, Donald H. General Purpose Software. In DATAMATION, Jan. 1968, p. 22.

Woolley, G. OVERALL VIEW OF DATA MANAGEMENT SYSTEM PROCESSING. Limited Note, Leasco Systems & Research Corp., Bethesda, Md.